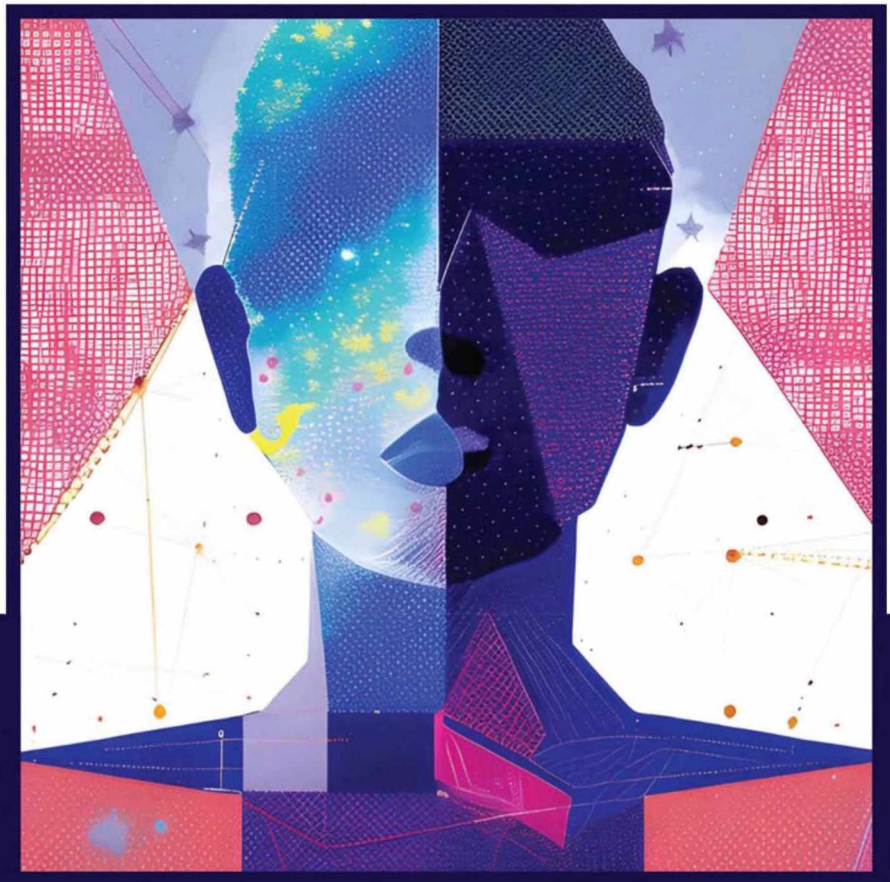
DEEPFAKE: A VALÓTLAN VALÓSÁG

Szerkesztette
Aczél Petra • Veszelszki Ágnes



MÉDIATUDOMÁNYI KÖNYVEK

DEEPFAKE:
A VALÓTLAN VALÓSÁG

Szerkesztette
Aczél Petra – Veszelszki Ágnes

Gondolat Kiadó
Budapest, 2023

A kötet megjelenését támogatta
a Nemzeti Média- és Hírközlési Hatóság Médiatanácsa.

www.gondolatkiado.hu
facebook.com/gondolat

A kiadásért felel Bácskai István
A borítóképet a Canva AI Image Generator programjával
Mezriczky Marcell készítette
A kötetet tervezte Lipót Éva

ISBN 978 963 556 460 6

Tartalom

JOG

INFLUENSZEREK

PSZICHOLÓGIA-PEDAGÓGIA

Köszöntő

A virtuális valóságban rejlő lehetőségek, a mesterséges intelligencia fejlődése olyan áttöréseket hozott, melyek néhány évtizeddel ezelőtt még a tudományos-fantasztikus kategóriába tartoztak. Robotok önálló tudattal, mások irányítása egy elképzelt világban, a valóságban nem is létező véleményvezérek követése a közösségi médiában. A sort napestig lehetne folytatni, hiszen mára igazzá vált az a szakállas reklámszlogen, miszerint a valóságnak csak a képzelet szabhat határt. Azonban, mint minden technológiai újítás, fejlesztés, számos kihívást és kockázatot is rejt magában.

Az emberiség elképesztő ütemben halad a technológiai változás útján. Olyan gyorsan, hogy a kivételesen gazdag és leleményes magyar nyelv nyomába sem tud érni. Szinte havonta kell újabb és újabb fogalmakat megtanulnunk. Még meg sem értettük teljesen, hogy mi a *blockchain* és a *bitcoin*, máris nyakunkon az *AI prompt engineering*. A *deepfake*-kel látszólag könnyebb a dolgunk, hiszen a képmanipuláció egyidős a fényképezéssel, sőt, ha Cyrano de Bergerac történetére gondolunk, még kép sem kell hozzá. A digitális kor egyik legizgalmasabb, ámde annál komolyabb problémája a deepfake, ami nem más, mint valamiféle mesterséges médiatermék, ami a *deep machine learning* technológia révén jön létre, és ami lehetővé teszi a tökéletesnek, teljesen hitelesnek tűnő, de manipulált videók, fényképek és hangfelvételek létrehozását.

A jelenség új korszakot nyitott az információmanipuláció terén, az álarc mögé bújtatott álhírek és megtévesztő tartalmak alapjaiban rendítik meg a valóságérzékelésünket, fenyegetik az információs integritást, és kételyeket ébresztenek bennünk. Egyre nehezebb felismerni a különbséget valós és virtuális között, a ránk gyakorolt hatásuk könnyen odavezethet, hogy elveszítjük a megbízható források azonosításának képességét. Ezen kihívások elkerülhetetlenné teszik a technológia alapos megismerését, annak működését és okozott hatásait.

Az adatbiztonság és a személyes magánélet védelme kiemelkedő fontosságú terület, mert a szakavatott támadók könnyedén manipulálhatják adatainkat és ezzel együtt egész személyazonosságunkat. Akár egy karrier, egy meghitt kapcsolat vagy egy élet kerülhet veszélybe az ilyen típusú visszaélések miatt. Ezért kell komoly figyelmet fordítani arra, hogy az online térben milyen információkat osztunk meg magunkról és családtagjainkról, vagy milyen forrásból tájékozódunk.

Kiemelt feladat, hogy fokozottan odafigyeljünk a deepfake-videók és hasonló tartalmak azonosítására és megjelölésére. Ez idővel egyre nehezebbé válik, mivel a technológia fejlődésével a hamisítások is egyre valószerűbbek lesznek. Ezért fontos, hogy az egyes online platformok és szolgáltatások olyan hatékony eszközök-

kel legyenek felszerelve, amelyek alkalmasak a manipulált tartalmak felderítésére. A felhasználók körében elsődleges szemponttá válik a digitális készségek fejlesztése és az információbiztonság iránti tudatosság.

Ugyanakkor az sem kívánatos, hogy egy jelenség sötét oldala túlságosan is beárnyékolja annak előnyeit és lehetőségeit. Kétségtelen, hogy a deepfake kreatív és szórakoztató célokra is használható. Sőt, a filmekben vagy videójátékokban ennek segítségével már olyan élethű látványvilágot és effekteket hoznak létre, amelyeket korábban elképzelni sem tudtunk. Játsszunk el a gondolattal, mekkora élmény lenne számunkra, ha a régi idők hírességei ismét feltűnnének a mozivásznon, Humphrey Bogart újra nyomozhatna, vagy Frank Sinatra újra koncertet adhatna a „Bűn városában". De segíthet a sérült, régi videók és hangfelvételek restaurálásában, így az általuk képviselt értékes történelmi emlékek még hosszú időre megőrizhetők és élvezhetők maradnak az elkövetkező generációk számára is.

Az eddig leírtakból világosan látszik, hogy a helyzet nem egyszerűen fehér vagy fekete, jó vagy rossz. Vannak előnyei és benne a végtelen lehetőség, de mind az egyén, mind pedig a társadalom szintjén tisztában kell lenni a veszélyeivel is. Az adatvédelem és a digitális biztonság mindannyiunk felelőssége. Az élet és a technológia közötti határvonalak egyre inkább elmosódnak, de törekednünk kell arra, hogy megőrizzük az egyensúlyt e két világ között. A valódi kapcsolatokat és élményeket nem lehet mesterséges szereplőkkel helyettesíteni. Az innováció és a technológia segít minket a fejlődésben, de tudnunk kell, hogy a valós érzelmek, a kézzelfogható dolgok és az összetartozás érzése tesz minket igazi és igazán emberré.

Bízunk benne, hogy ez a könyv, amely szakterületük elismert és feltörekvő fiatal kutatóinak írásait gyűjti össze, segít jobban belelátni ebbe a világba, megérteni a folyamatokat, és kellőképpen felvértezi az olvasót azzal a tudással, amely segítségével felismerheti és elkerülheti a manipulált tartalmak által létrehozott csapdákat és buktatókat. Legyünk nyitottak az új ismeretek befogadására és elsajátítására, mert csak így készülhetünk fel kellőképpen a digitális kor kihívásaira!

Koltay András – Szadai Károly

Előszó

Nézzük a videófilmet, ismerős a szereplő, akit látunk. Hallgatjuk, amit mond, és arra gondolunk, nem hittük volna, hogy egyszer ilyen dolgok hagyhatják el a száját. Meg sem fordul a fejünkben, hogy ő sem hitte volna. Egész pontosan az a politikus, művész, színész, médiaszereplő orvos vagy tanár sem, akinek éppen szimulált változatával találkoztunk. Feltehetően ő sem számolt azzal, hogy egyszer nyilvánosságba lépő szintetikus mására és mondandójára fog rácsodálkozni – rémülten vagy éppen elégedetten.

A deepfake korunk imposztora. Olyan tartalom, kép, üzenet, amely mögött a gyorsan tanuló mesterséges intelligencia dolgozik azért, hogy egy, a valóságra megtévesztésig hasonlító, valójában azonban nem létező médiaeseményt hozzon létre. Például az amerikai elnök egy soha el nem hangzott beszédét. Háborúban álló felek vezetőinek megadó vagy felbujtó üzeneteit. Egy régen elhunyt művészt, aki ma „személyesen" fogadja a saját kiállítására érkező vendégeket. Vagy a szintén régen halott színész 2023-as, új szerepjátékát. A deepfake olyan médiatartalom, amellyel ma már bárhol találkozhatunk, anélkül megszokva jelenlétét, hogy többet és mélyebben gondolkodnánk róla. Tanulmánykötetünk az utóbbiban, a deepfake pontosabb megismerésében támogatja az Olvasót. Magyar nyelven és a régióban elsőként olyan írások összefoglaló könyve ez, amely a deepfake jelenség egészéről és részleteiről is naprakész képet ad. Az írásokat jegyző szerzők saját szakterületük avatott és ismert hazai kutatói, e kötetben pedig mindannyian, újszerűen, a deepfake témája felé konvergálják nézőpontjaikat, kutatási eredményeiket.

A tizennégy tanulmány öt fejezetre tagolja a kötetet. Az első a deepfake kommunikációs vetületeit dolgozza fel. Veszelszki Ágnes olyan általános, az előnyöket, hátrányokat is bemutató képet ad a deepfake-jelenségről, amely nélkül a későbbi részletek nehezen lennének érthetők, Aczél Petra a deepfake mint hazug üzenet interakciós jóváhagyását elemzi, Mezriczky Marcell pedig a deepfake-kel kapcsolatos hazai, fél évtizedes médiareprezentációt vizsgálja tanulmányában. A második, technológiafókuszú fejezet a deepfake programozási jellemzőibe, várható jövőjébe, kiberbiztonsági kockázataiba nyújt betekintést. Keleti Arthur a deepfake és a hitelesség emberi-technológiai-informatikai összefüggéseit járja körül, míg Szabados Levente a deepfake-technológia múltját és közeli jövőjét rajzolja elénk. A fejezetegység fontos zárásaként Krasznay Csaba írása a deepfake hatását tárgyalja a kiberbűnözésben, kiberkémkedésben és kiberhadviselésben. A kötet harmadik nagy egysége a jogi aspektusokra világít rá, sokirányú betekintést nyújtva.

Eszteri Dániel a deepfake adatvédelmi értékelését taglalja, Lendvai Gergely Ferenc a deepfake és a szólásszabadság sajátos viszonyát dolgozza fel, zárásként Miklós Gellért a deepfake-tartalmak európai szabályozását elemzi. Az influenszerek, a tartalom- és ismertségipar adja a kötet negyedik nagy szakaszát. Itt találjuk Gocsál Ákosnak a manipulált beszédről, Guld Ádámnak a deepfake-influenszerek jellemzőiről, karrierjéről, illetve Horváth Evelinnek a deepfake szépségideálhoz való kapcsolódásáról szóló tanulmányát. Az ötödik, egyszersmind zárómodul a deepfake pszichológiai és pedagógiai vetületeit tárja fel. Tari Annamária a manipulált videók pszichológiai hatásairól, Kárpáti Andrea pedig a deepfake oktatási jelenéről, jövőjéről, a legfontosabb kihívásokról ír munkájában.

A tanulmánykötetben tizennégy különféle aspektus, sokféle tudományterület beszél tehát ugyanarról a jelenségről: a deepfake-ről, amelynek alkotói és befogadói, ragadozói és áldozatai, gazdagjai és szegényei egyaránt lehetünk. Abban bízunk, hogy a kötet elolvasása után kinek-kinek könnyebb lesz döntenie, melyik oldalt választja.

Budapest, 2023 májusában

Aczél Petra és Veszelszki Ágnes

KOMMUNIKÁCIÓ

VESZELSZKI ÁGNES

Deepfake: kételkedés a kételyben

A fejezet három (egy-egy névutópárral jellemezhető) dilemmát vizsgál meg a deepfake-kel, illetve az ennek alapjául szolgáló képfelismerő és -alkotó mesterséges intelligenciával kapcsolatban. Az első az elől – felé *névutókkal jellemezhető skála, amelynek egyik oldalán az automatikus kép- és személyfelismerés negatív következményei (mint az azzal összekapcsolt szociális értékelőrendszer, illetve az előítéletes algoritmus), a másik felén pedig a pozitívumai (kiemelten a szociális biztonság) állnak. A második,* mellett – ellen *névutóval leírható dilemma a deepfake használatának előnyeit (az üzletben, a videótechnikában, a szórakoztatásban) sorakoztatja fel, a hátrányainak és veszélyeinek mérlegelésével (többek között: az ideálisnak tartott testkép megváltozása, a kiberbűnözés új formái, a hitelesség fogalmának újraértelmezése). Kifejezetten ez utóbbihoz, a hitelesség kérdéséhez kötődik a harmadik, az* alapján – nélkül *névutókkal jelzett dilemma, amely a vizuális bizonyítékok politikáját, illetve a kétely kételye paradoxont mutatja be.*

Kulcsszavak: képalkotó és képfelismerő algoritmus, videómanipuláció, szövegtenyésztés, hitelesség

1. PÁNIK ÉS DILEMMÁK

Elérkezett a stockfotó-adatbázisok vége, megszűnnek a grafikusi állások, hiszen itt van a mesterséges intelligencia korszaka, amelyben már a szemünknek sem hihetünk? A mesterségesintelligencia-fejlesztések mai űrversenye az előbbi háromhoz hasonló, újfajta aggodalmakat hoz magával. Ugyanilyen pánikreakció volt megfigyelhető a deepfake-technológia első, 2017-es megjelenésekor: az *MIT Technology Review* szerzője azt feltételezte, hogy a mesterséges intelligencia száz évvel is visszavethet minket abban, ahogyan híreket fogyasztunk („AI Could Send Us Back 100 Years When It Comes to How We Consume News", Snow 2017), a *The Atlantic* a valóság összeomlásáról írt („collapse of reality", Foer 2018), és a *The New Yorker* feltette azt a kérdést, hogy az AI korában hihetünk-e még a szemünknek („In the age of A.I., is seeing still believing?", Rothman 2018).

E fejezetben három névutópárral jellemezhető dilemmát vizsgálok meg a deepfake-kel, illetve az ennek alapjául szolgáló képfelismerő és -alkotó mester-

séges intelligenciával kapcsolatban. Az első az *elől – felé* névutókkal jellemezhető skála, amelynek egyik oldalán az automatikus kép- és személyfelismerés negatív következményei (mint az azzal összekapcsolt szociális értékelőrendszer, illetve az előítéletes algoritmus), a másik felén pedig a pozitívumai (kiemelten a szociális biztonság) állnak. A *mellett – ellen* névutókkal jellemezhető második dilemma a deepfake használatának előnyeit (az üzletben, a videótechnikában, a szórakoztatásban) sorakoztatja fel, a hátrányainak és veszélyeinek mérlegelésével (többek között: az ideálisnak tartott testkép megváltozása, a kiberbűnözés új formái, a hitelesség fogalmának újraértelmezése). Kifejezetten ez utóbbihoz, a hitelesség kérdéséhez kötődik a harmadik, az *alapján – nélkül* névutókkal jelzett dilemma, amelynél a vizuális bizonyítékok politikáját, illetve a kétely kételye paradoxont mutatom be.

2. A DEEPFAKE KERETEI

A mesterséges intelligencia két szintjéből – 1. a szűk, alkalmazott vagy gyenge MI, amely bizonyos folyamatokban elérheti, akár felül is múlhatja az emberi teljesítményt, ám csupán abban a tevékenységben, amelyre tervezték (például: adatelemzés, beszédfelismerés és -szintetizálás, képfelismerés és -módosítás); 2. az erős vagy emberi szintű általános mesterséges intelligencia *(artificial general intelligence, AGI),* amely képessé válhat az önálló gondolkodásra és cselekvésre, komplex kérdésekben is – jelenleg az első fázisban tart a technológiai fejlesztés (Özkiziltan–Hassel 2021; OECD 2019 nyomán Mezriczky 2021). A mesterséges intelligencia ugyan képes más eszközökkel feldolgozhatatlan mennyiségű adat elemzésére, komplex használatára, ám egyelőre emberi irányítással, bizonyos korlátozott területeken.

Ennek kapcsán azonban megfontolandó az a technikai-filozófiai fordulat, amelyre Chris Bishop, a Microsoft kutatási vezetője hívta fel a figyelmet: a számítógépek történetének első negyven évében programoztuk a számítógépeket, a következő negyven évben tanítani, trenírozni fogjuk őket (Heaven 2021). Korábban a beszéd- vagy képfelismeréshez a programozók szabályokat adtak meg a számítógépnek – ezzel szemben az informatikusok a gépi tanulás korszakában már neurális hálózatokat alkotnak, amelyek saját maguknak írnak szabályokat. Ráadásul ez utóbbi esetén nem kell közvetlenül parancsokat beírni vagy akár gombokat megnyomni: az egyes műveletek végrehajtásához nem szükséges billentyűzet vagy képernyő az emberrel való interakcióhoz. Ez alapvetően más, újfajta gondolkodási módot jelent – állítja Heaven (2021).

A mesterséges intelligencia már jelenleg is képes arra, hogy egy videóban a neurális hálózat és a fellelhető információk alapján egy szereplő arcát egy másikéra cserélje, az eredeti szereplő hangját utánozza, oly módon, hogy az arc- és hang-

csere a hétköznapi felhasználó számára felismerhetetlen legyen, tehát deepfake-et állítson elő.

Mezriczky Marcell (2021) hat különböző deepfake-definíció összehasonlítása révén alkotta meg a saját fogalommagyarázatát, miszerint a deepfake-videó olyan, „a mesterséges intelligencia által létrehozott [...] mozgóképes tartalom, amelyben a tartalom eredeti szereplőjének képmását és/vagy hangját egy, tőle idegen személy képmásával és/vagy hangjával helyettesítik, valamely kívánt hatás elérése érdekében". Működési elv szerint több deepfake-típus különíthető el egymástól: például az arccsere *(face swapping);* az arc újraanimálása *(face reanimation);* a tárgyeltüntetés *(object removal);* valamint a teljesen mesterségesen létrehozott kép és hang (*CGF, computer-generated face; speech synthesis, text-to-speech;* vö. Horváth–Mezriczky 2021).

A deepfake összegezve tehát digitálismédia-manipulációt jelent: ultrarealisztikus, gépi tanulással létrehozott hamis(ított) videót, amelynek a szereplői olyan dolgokat tehetnek vagy mondhatnak, amit nagy valószínűséggel nem tennének vagy mondanának (Dobber et al. 2020: 1; Veszelszki 2021: 97). Valós képi és hangzó anyag felhasználásával az ún. neurális hálózat olyan videószekvenciák létrehozására képes, amelyek alkalmasak lehetnek az emberek megtévesztésére. A deepfake abban különbözik a photoshoppolt képektől, hogy nem csupán a szemre, hanem a fülre is hat. A gyanútlan és gyakorlatlan videónézőt az egyre jobb és jobb minőségben előállított deepfake-videók könnyen becsapják. Ugyancsak megnehezíti a deepfake felismerését, hogy ezek a videók nagyrészt valódi képsorozatok, a deepfake-előállítók csupán kisebb részeket (például arckifejezést, hangot) változtatnak meg rajtuk (Dobber et al. 2020: 2). Ily módon a tényeket és a fikciót még nehezebb elkülöníteni egymástól.

3. HÁROM DILEMMA

A deepfake jelenség a jogi, etikai, pszichológiai-pedagógiai, kiberbiztonsági, informatikai stb. területek mellett a kommunikációtudományi diskurzusra is erőteljesen hat. A vonatkozó kommunikációs vizsgálatokat foglalom össze a következőkben három fő témakör kapcsán: elsőként a deepfake létrehozásához szükséges három alapvető technológia – kép- és arcfelismerés, hangklónozás, gépi képalkotás – előnyeit és hátrányait vetem össze (ehhez kapcsolom az *elől* és a *felé* névutókat). Másodjára a deepfake-technológia világos és sötét oldalát, előnyeit és közvetlen veszélyeit taglalom (ahol a *mellett* és az *ellen* lesz az iránymutató névutópár). Végül az információmanipuláció közvetett hatásait jellemzem az eddigi kutatások felhasználásával (a témához illeszkedő névutók: az *alapján* és a *nélkül*). Az itt bemutatott három, névutópárokkal kiemelt téma egyúttal három, skálaszerű dilemmát is jelöl, amelyek közötti választás kommunikációs szempontokat is alkalmazó mérlegelés tárgya lehet.

3.1. Elől – felé

A minél hatékonyabb, azaz megtévesztőbb deepfake létrehozásához alapvetően három fő fejlesztési irány járul hozzá: a kép- és arcfelismerés, a hangklónozás, illetve a gépi képalkotás. A következőkben ezeket vesszük sorra.

3.1.1. Kép- és arcfelismerés

A – *track, trace, monitor* fogalomhármassal jellemezhető – digitális megfigyelés fejlesztése során a mesterséges intelligenciát az osztályozás és a címkézés *(classification & labelling system)* „képességére" trenírozzák (Lee 2020). Ez a technika is hozzájárul a deepfake létrehozásához, egyúttal – mint az alábbiakból kitűnik – a digitális megfigyeltség új formáit is magával hozza, és megváltoztatja a beleegyezés jogának normáit (Hao 2021).

Az arcfeldolgozó technológia *(facial processing technology, FPT)* egyre inkább áthatja az életünket számos területen: alkalmazzák már többek között köztereken, lakóparkokban, üzletekben, repülőtereken, iskolákban, koncerteken (Raji–Fried 2021). Egy amerikai kormányzati megfigyelési szakértő, Steven Feldstein szerint 2012 és 2020 között a vizsgált 179 országból 77 használta az AI-alapú megfigyelést, 61 ország pedig a digitális arcfelismerést (Johnson 2023). Tudósítások szerint a technológiát legújabban Iránban arra is alkalmazzák, hogy a hidzsábot a jogszabály ellenére nem viselő nőket ezzel azonosíthassák, és az adatbázis alapján megbüntethessék őket (Johnson 2023). A hivatalos magyarázat szerint a kamerák és az arcfelismerő rendszerek használatával csökkenthető az utcákon a rendőri jelenlét (és ezáltal a civilek és a rendőrök között az összecsapások száma is; Johnson 2023).

A Facebook fejlesztői által 2014-ben bevezetett Deepface modell volt az első, mélytanulással működő – egyébként a Facebook profilfotóin képzett – arcfelismerő rendszer (Raji–Fried 2021: 3). Bemutatásakor a siker átütő volt: a 97,35%-os pontosságával 27%-kal jobb volt, mint bármely korábbi arcfelismerő technika. Ekkortól vált a neurális hálózatok használata az arcfelismerésben általános módszerré (Raji–Fried 2021: 3), amelyben a személyazonosítás – az arckép és azonosító adatainak összekapcsolása – helyett a kategorizációra, osztályozásra került a hangsúly (Hao 2021). Ennek kapcsán gyakran merül fel a rasszista, elfogult algoritmus problémája (vö. Feathers 2020), illetve – az AI-modellek növekvő adatigénye miatt – az adatterjesztésnek, a képek felhasználásához való hozzájárulásnak, a magánélet védelmének a kérdése (Raji–Fried 2021: 8).

3.1.2. Hangklónozás

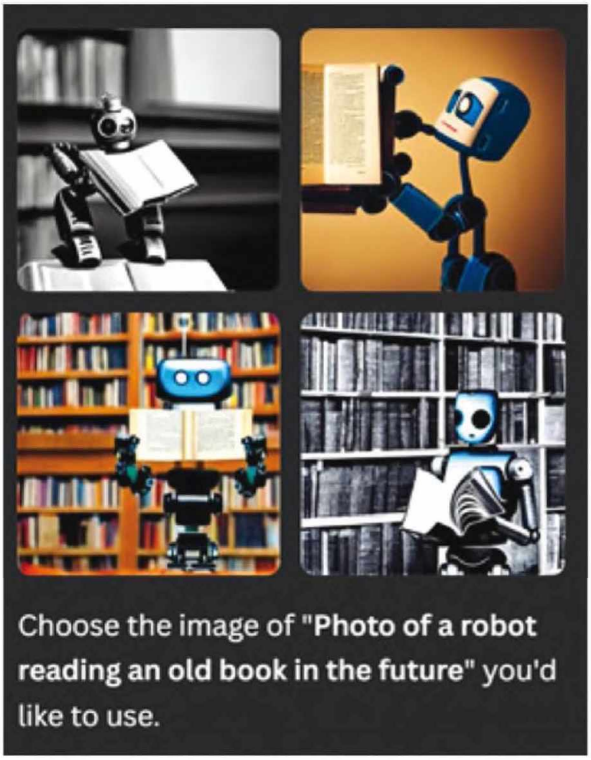
A mesterséges intelligencia által létrehozott hang ugyancsak jelen van a mindennapjainkban: többek között az otthoni és virtuális asszisztensek (mint a Siri vagy Alexa) is a természetes nyelvfeldolgozásra *(natural language processing, NLP)* építenek.

Emellett a hanggeneráló és -klónozó alkalmazásokat is folyamatosan fejlesztik. Például a Google kutatói által létrehozott AudioLM (Borsos et al. 2022) képes arra, hogy néhány másodpercnyi hanganyag alapján természetes hangzású beszédet és zenét alkosson, megadott stílusban, az eredeti felvételtől szinte megkülönböztethetetlen minőségben, akár olyan összetett hangok esetén is, mint a zongorajáték vagy az emberi beszéd (Xu 2022). A technológia abban különbözik más, korábbi hangfeldolgozó rendszerektől, hogy nem igényel transzkripciót és címkézést, ezáltal sokkal gyorsabban, szinte emberi közreműködés nélkül tud működni. A ChatGPT-3-hoz hasonló nyelvi modellek elvéhez hasonlóan (amelyek megjósolják, hogy jellemzően milyen mondatok és szavak követik egymást) a néhány másodpercnyi hang betáplálását követően a rendszer megjósolja a következő hangokat (Xu 2022). Nem csupán zenei hangok generálására korlátozódik a tevékenysége: eredeti írott szöveg nélkül (vagyis nem a *text-to-speech* technológia használatával) emberi beszéd generálására is képes, az eredeti beszélőéhez hasonló akcentus, hangsúly és beszédstílus utánzásával (ám egyelőre helyenként szemantikailag nem koherens, értelmetlen beszédproduktumok megalkotásával; Borsos et al. 2022). Mivel a rendszer az adatbázis alapján megtanulja, milyen típusú hangrészletek fordulnak elő gyakran együtt, így képes visszaadni a beszélt nyelv sajátos, írásban esetleg nem jelölt szupraszegmentális jegyeit is (Xu 2022).

A zene- és hanggeneráló rendszer használható többek között videók természetes aláfestő zenéjének automatikus megalkotására, fogyatékossággal élők különböző internetes eszközökhöz való hozzáférésének elősegítésére, egészségügyi robotok fejlesztésére. Azonban a technológia új etikai és jogi kérdéseket is felvet: például a rendszer tréningezésére használt zenék, hangok alkotói kapnak-e az AI alkotta produktumokból jogdíjat; illetve a valóságostól megkülönböztethetetlen beszéd révén az eszköz hozzájárulhat félrevezető, megtévesztő információk terjesztéséhez (*spoofing*, rossz szándékú megszemélyesítés; Xu 2022). Ez utóbbi miatt merült fel az a javaslat, hogy a természetes hangoktól való megkülönböztetés érdekében a mesterséges intelligencia által generált produktumokba építsenek be audiovízjeleket.

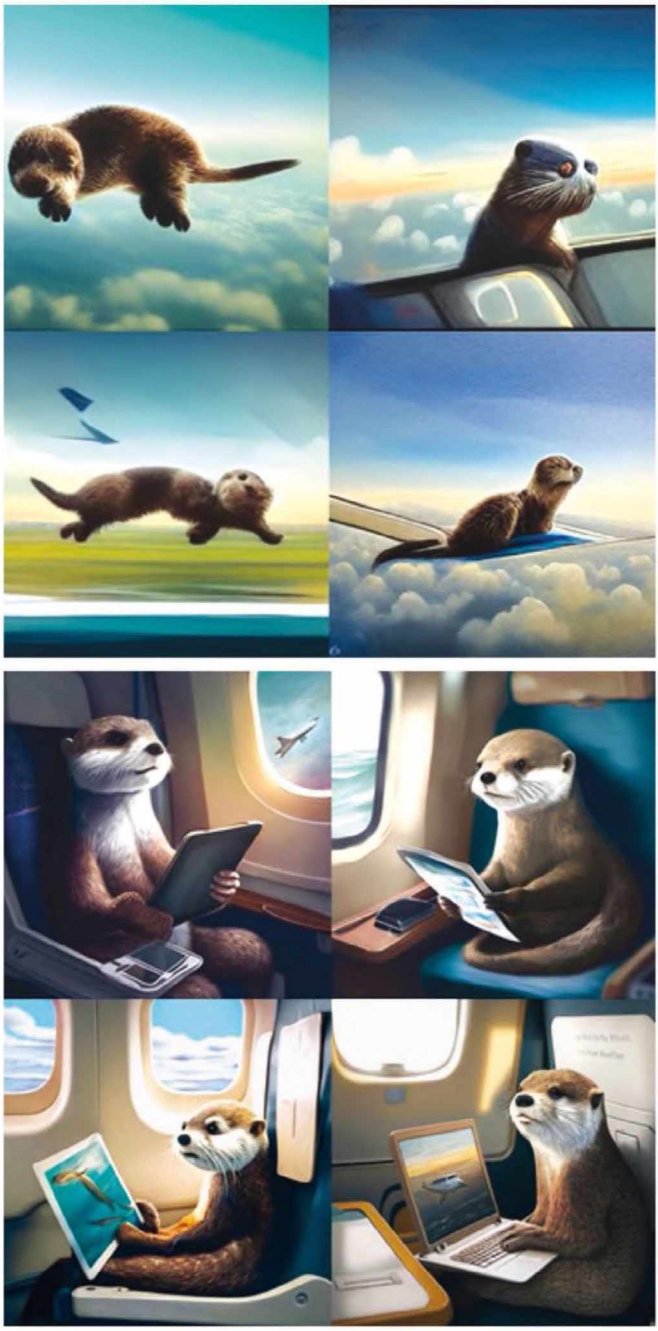
3.1.3. Gépi képalkotás

A mesterséges intelligenciára építő, szövegből képet alkotó megoldások (*AI art generator*, vö. DALL-E 2, Midjourney, Stable Diffusion, Imagen: rendre W1, W2, W3, W4) már a hétköznapi felhasználók számára is szinte különösebb előképzettség nélkül elérhetők (1. ábra).



1. ábra. Az „egy robot régi könyvet olvas a jövőben" prompttal készült képek a Canva segítségével (forrás: saját szerkesztés a Canva felületén)

A *text-to-image*, azaz szövegből képet alkotó alkalmazások fejlesztése 2022 őszén a korábbihoz képest is magasabb fordulatszámra kapcsolt. Erre a fejlesztési sebességre mutat egy példát Ethan Mollick (2022): az ugyanazzal a szöveges utasítással, ún. prompttal egy hónap eltéréssel sokkal pontosabb, részletesebb, „valósághűbb" képet volt képes megalkotni a Midjourney alkalmazás harmadik, illetve negyedik verziója (2. ábra).



2. ábra. A képalkotó alkalmazások fejlődése: az „otter on a plane using wifi” 'egy vidra wifit használ egy repülőn' prompt eredménye egy hónap eltéréssel (Midjourney v3 vs. v4, forrás: Mollick 2022)

Milyen kérdéseket, esetleg aggályokat vethet fel a szövegből képet alkotó mesterséges intelligencia? Átalakítja, de legalábbis módosítja a művészetről alkotott elképzeléseinket: a milliárdnyi felhasznált adatból az AI-alapú képgenerátorok új, korábban nem létezett vizuális terméket hoznak létre – kérdés lehet, hogy ez vajon kreatív alkotásnak számít-e, vagy csupán meglévő művek újrahasznosítása. A közeljövőben hatni fog a grafikusi, vizuális gyakorlatokra (Nyiri 2022): akár az AI-képgenerálás tudatos elkerülésével, tiltásával, akár a használatának megtanulása (a promptírás, azaz a megfelelő szöveges utasítások megadása újfajta tudást jelöl) és az AI generálta képek-videók tudatos alkalmazása, szerkesztése révén. A mesterséges intelligencia alkotta vizuális termékeket (egyelőre) nem védi szerzői jog – illetve a felhasznált művek alkotói (már csak a visszakövethetetlenség okán) sem kapnak jogdíjat.

További etikai problémák merülnek fel az AI képgenerátorok kapcsán, és erre nagyon szemléletesen mutat rá az, hogy a legnagyobb, szövegből képet alkotó alkalmazások oldalán nem vagy alig lehet emberi arcokat találni: ehelyett antropomorfizált, kedves állatfigurákat láthatunk tipikusan emberi tevékenységeket végezni.[1] Will Douglas Heaven (2022b) szerint ennek két oka is van: az aranyos állatok jelenlétének ebben a körben egyrészt van egy technikai, másrészt pedig egy kommunikációs, PR-jellegű magyarázata. Technikai oknak számít, hogy ha nagyon különböző, távoli koncepciókat (mint amilyen a vidra, a repülő és a wifihasználat) kell az algoritmusnak összekapcsolnia, akkor sokkal pontosabban megtanulja, hogy miként tud ebből viszonylag értelmes alkotást létrehozni. Azonban az, hogy mind az OpenAI DALL-E 2, mind a Google Imagen nyilvános felületén majdnem kizárólag kedves, bájos, a promptoknak szinte tökéletesen megfelelő képeket találhatunk, egy komoly etikai nehézségre mutat rá. Ezek a gondosan kiválogatott képek elrejtik a gyűlölködő, ártalmas sztereotípiákra építő produktumokat – pedig a rendszerek készítéséhez használt webes adatbázisok alapján valószínűsíthető, hogy ezek is megjelennének. Az Imagen weboldalán a társadalmi hatások leírásakor az AI ilyen jellegű etikai kihívásaira explicit módon is reflektálnak.[2]

[1] A „This person does not exist" oldal portrékat hoz létre nem létező emberekről, az AI segítségével, de szöveges input nélkül; vö. W5.

[2] [1] „…the data requirements of text-to-image models have led researchers to rely heavily on large, mostly uncurated, web-scraped datasets. While this approach has enabled rapid algorithmic advances in recent years, datasets of this nature often reflect social stereotypes, oppressive viewpoints, and derogatory, or otherwise harmful, associations to marginalized identity groups. While a subset of our training data was filtered to removed noise and undesirable content, such as pornographic imagery and toxic language, we also utilized LAION-400M dataset which is known to contain a wide range of inappropriate content including pornographic imagery, racist slurs, and harmful social stereotypes. Imagen relies on text encoders trained on uncurated web-scale data, and thus inherits the social biases and limitations of large language models. As such, there is a risk that Imagen has encoded harmful stereotypes and represen-

Erre a megoldás egyelőre az, hogy csak bizonyos használók férhetnek hozzá az eszközhöz (DALL-E 2), vagy egyáltalán nem nyilvános a felhasználói felülete (Imagen).

3.2. Mellett – ellen

3.2.1. A deepfake sötét oldala

Az AI közreműködésével megalkotott, a kiterjesztett valóság részeként működő filterek is egyfajta deepfake-nek tekinthetők. A közösségi médiában használható filterek képesek a szereplő arcának és testének valós időben történő „megszépítésére" (a jelenlegi női szépségideál szerint nagy szemek, kis orr, húsos száj, dús haj, hibátlan bőr kialakítására a digitális képen), illetve fantáziajegyek (például kutyafülek, ördögszarvak, lila szemek) hozzáadására is. 2019 októberében a Facebook a potenciális negatív hatások miatt betiltotta a torzító filterek használatát, majd 2020 augusztusában ismét engedélyezte ezeket, a tiltást csak a plasztikai sebészetet népszerűsítő filterekre fenntartva (Ryan-Mosley 2021). 2021 nyarán jelent meg az a jogszabály(tervezet) Norvégiában, miszerint az influenszereknek fel kell tüntetniük, ha a hirdetett posztjukban szereplő képet digitálisan szerkesztették (Grant 2021). A kormányzat célja a rendelkezéssel az, hogy általa az idealizált, ám egészségtelen és valószerűtlen test- és arcképek közösségi médiában történő megjelenése miatt a fiatalokra nehezedő nyomást, a testképzavarral összefüggő mentális problémákat csökkentsék. A posztban a terv szerint azokat a retusálási és képmanipulációs módszereket kell jelezni – hasonlóképpen a natív és a hirdetési tartalom megkülönböztetéséhez –, amelyekben a reklámban szereplő személy teste (testalkata, mérete, bőre) eltér a valóságostól. A rendelettervezet lefedi a filterek és a digitális módosítások használatát is. Az, hogy a jogi szabályozás területét is elérte a közösségi médiában megjelenő manipulált képi tartalmak kérdése, jelzi a témakör jelentőségét. A testképre kifejtett hatás *(body dysmorphia)* mellett a közösségi oldalak által a filterek használata révén begyűjtött biometrikus adatok sokasága is aggasztó lehet (Ryan-Mosley 2021).

A deepfake a visszaélések esetén visszatérhet a kiindulópontjához, ahol először megjelent, mégpedig a pornóhoz – figyelmeztetnek a szakértők (vö. Snow 2022). Az első, a főként a Redditen publikált, amatőr deepfake-videók pornófelvételekben jelenítették meg az arccsere segítségével ismert színésznők arcát (Gal Gadot, Scarlett Johansson), miközben ők valójában nem szerepeltek ezekben a felnőttfilmekben. Az ilyen jellegű tartalmat a Reddit már 2018-ban letiltotta, később a Pornhub nevű pornómegosztó oldal ugyanígy tett a deepfake-pornóval. Az újabb

tations, which guides our decision to not release Imagen for public use without further safeguards in place" (forrás: W4).

fejlesztések azonban már nemcsak az arccserével „hamisított", hanem a teljes mértékben a minták alapján generált, de új tartalomként megjelenő pornográf képek megalkotását is lehetővé teszik (és minden bizonnyal az AI generálta felnőttvideók is hamarosan piacra kerülnek).

Az AI nem pornótartalmak esetén is képes erotikusabb kinézetet rendelni a felhasználókhoz. A 2022 végén különösen népszerűvé vált Lensa nevű AI-alapú applikáció az előfizetők által feltöltött portrékból készít „varázslatos avatárokat". Bár a felhasználói szabályzat szerint nem szabad sem meztelen, sem gyermekekről készült képet feltölteni, egy kísérlet szerint (Snow 2022) egyik szabály megsértését sem szankcionálja az applikáció ténylegesen.[3] Ráadásul az app zavarba ejtő mértékben szexualizálja az alanyokat, különösen a nőket (nagyobb melleket rendel az avatárokhoz, erotikus pózokban, akár meztelenül ábrázolja őket; Snow 2022). A rendszer a trenírozásához használt adatok alapján mintákat azonosít, és az AI generálta végeredmény gyakorlatilag szűretlenül jelenik meg.

A deepfake a testképre gyakorolt hatás mellett széles körű jogi-morális problémákat is fölvet: újfajta – ember vagy akár robot kontrollálta – csaló, megtévesztő tevékenységekre kell felkészülni. A deepfake az elemzések szerint a közeljövőben nagy szerepet fog játszani a magán- és közéleti bosszúvideók előállításában (vö. bosszúpornó), különböző bűnügyek feltárásában valódi vagy hamis bizonyítékként, továbbá kiberbiztonsági aggályok merülhetnek fel általa (Harwell 2020; Ajder et al. 2019; Öhman 2020).

A test fokozott digitalizációja révén (például különböző felületeken hangazonosító megadása, publikus közösségimédia-profilok használatával) viszonylag könnyedén utánozható bárki digitális identitása. A biometrikus hangazonosítást veszélyeztetheti a hangutánzás *(spoofing)*, amely által a csalók szenzitív információkhoz férhetnek hozzá, beléphetnek az adott személy különböző felhasználói fiókjaiba. Az auditív deepfake-kel szenzitív információ (hitelkártya-adatok, jelszó) is kicsalható a kiszemelt személytől *(phishing)* – amint ez meg is történt egy brit energiacég ügyvezetőjével, aki azt hitte, hogy az anyacég német vezetőjével beszél telefonon, ami miatt 220 000 eurónyi összeget utalt át egy (egyébként magyar) bankszámlára (Damiani 2019). Később kiderült, hogy a csalók AI-alapú hanggenerátort használtak, amely olyan megtévesztő volt, hogy még a cégvezető németes akcentusát, jellegzetes beszéddallamát is képes volt utánozni. A csalók magabiztosságára utal, hogy háromszor is telefonáltak, de csupán az első átutalást hajtotta végre az ügyvezető (az összeg a magyar számláról mexikói, majd ismeretlen számlákra került), mind ez idáig a tetteseket nem azonosították.

Ugyanígy megjelenhet az ún. unokázós csalás *(grandparent scam)* AI-alapú változata: ennek során a csalók idős embereket hívnak fel, hogy az unokájuk bajba

[3] Megjegyzendő, hogy a felhasználók által feltöltött arcképek felhasználása, adatkezelése sem teljesen világos az app esetén.

került, és segítsenek neki egy, az áldozathoz küldött személyen keresztül a neki átadott pénzzel, értéktárgyakkal (Huffman 2020). Ez a csalási forma lehet még szofisztikáltabb a deepfake-kel, hiszen a segélykérés az unoka valódinak ható, utánzott hangján is elhangozhat (a rokoni viszonyokat márpedig könnyű feltárni a közösségi média segítségével).

További bűncselekmények is elkövethetők a deepfake használatával: hamisított kínos tartalom eltüntetéséért pénzt követelhetnek rágalmazók, továbbá a deepfake a cyberbullying, vagyis az online zaklatás alapjává is válhat. Mindezek mellett a bizonyítékok valóságtartalmára is kihatással van a deepfake: a bírósági eljárások, bűnügyi perek során azt is ellenőrizni kell majd, a bizonyítékként felmutatott kép-, hang- és videófelvétel nem hamisítás eredménye-e.

Az álhírek és dezinformáció témakörében egyre több szó esik a deepfake politikai szándékú felhasználásáról is. A technika már rendelkezésre áll ahhoz, hogy bármely politikai szereplő megpróbálja deepfake-kel hitelteleníteni, lejáratni az ellenfelét – vagy a saját érdekében politikai botrányt robbantson ki. A politikai befolyásolás megnyilvánulhat hamis kampánytámogatások gyűjtésében, közszereplők megrágalmazásában is. Mindez kihathat a demokratikus intézmények működésére, illetve a különböző politikai aktorok hatalmára (Bennett–Livingston 2018; Flynn et al. 2017; Bradshaw–Howard 2018), továbbá súlyosabb társadalmi konfliktusokhoz is vezethet. Dobber és kutatótársai (2020) egy, a politikai attitűdök deepfake-kel történő megváltoztatására irányuló online kísérletükben azt találták, hogy az általuk létrehozott deepfake-videó képes volt egy politikai szereplő megítélését megváltoztatni. A 278 vizsgálati alany válaszai alapján a politikus kedveltsége szignifikánsan csökkent az őt lejáratő, hamis videó megnézése után – ám a politikus pártja iránti elkötelezettség változatlan maradt a kontrollállapothoz képest. A politikai színtér mellett a deepfake-kel előállított tartalmak a gazdaság befolyásolására is képesek (például egy tőzsdei hír elterjesztésével).

3.2.2. A deepfake világos oldala

Természetesen nem csupán negatív felhasználási módokat ismerünk. Amivel az átlagos felhasználó leggyakrabban találkozik a közösségi médiában a deepfake kapcsán: azok a szórakoztató szándékú, paródia- vagy zenés videók, amelyekben az eredeti szereplő arca helyére a saját vagy bárki más arcát tudja a felhasználó egy applikáció használatával beilleszteni. A deepfake-humor, a deepfake-kel készült paródiák egyik lényeges eleme az esetlenség, a manipuláció könnyű felismerhetősége (Hao 2020b; Sholihyn 2020).

Mindezek mellett kísérleteznek a deepfake tudománykommunikációs, oktatási felhasználási lehetőségeivel is: egyfajta virtuális oktatóként hallható előadás Einsteintől a relativitáselméletről vagy éppen József Attilával is „elszavaltatható" az egyik költeménye, és lehet ezáltal a dinamikus storytelling része a deepfake-tartalom. Akár már nem élő emberek – rokonok – régi fotográfiáit lehet mozgóképpé tenni

a Deep Nostalgia program segítségével (a MyHeritage családfa-felkutató oldalon): ezáltal mosolyra fakadhat, kacsinthat valamely ősünk vagy akár egy híres személy is. A floridai St. Petersburgban található Dalí Museum deepfake-et használt a művész „megtestesítéséhez": Dalí köszönti a múzeumba látogatókat, megosztja velük a gondolatait a művészetről, és még szelfizhetnek is vele a látogatók (Martin é. n.), bár a deepfake elnevezést a fogalom negatív konnotációja miatt a múzeum igyekszik kerülni (Wilson 2019). 6000 kép, 1000 órányi gépi tanulás, több ezer oldalnyi kézirat szolgáltatta az interaktív Dalí-élmény alapját (Wilson 2019). A múzeum egy másik projektjében is épít a mesterséges intelligenciára: a DALL-E[4] segítségével műalkotást készít a látogatók álmaiból: a szövegként megadott álmot más látogatók szöveges utasításaival együtt a rendszer digitális, kivetítőn látható, „kollektív műalkotássá" formálja (Bicsérdi-Fülöp 2022). A Dalihoz hasonló realisztikus virtuális avatárok megtestesíthetnek történelmi szereplőket, elhunyt személyeket; de lehetnek akár virtuális popsztárok vagy tévébemondók is (vö. CGI-influenszerek). A hasonló projektekhez csak elegendő input (képi, hang- és/vagy mozgóképes tartalom) szükséges az erre a célra használt mesterséges intelligencia számára.

Az orvostudományban és rehabilitációban is segíthet a deepfake és a hozzá kapcsolódó hangszintetizálás: a hangjukat balesetben vagy betegségben elvesztett betegek szólalhatnak meg általa a saját (az eredetiről készült hangfelvételek alapján generált) hangjukon (Lee–Fung 2022; Martin é. n.).

A hangszintetizálás minden bizonnyal hatással lesz a filmszinkronpiacra is: a színészek eredeti hangján szinkronizálhatóak filmek, sorozatok – gyakorlatilag a hangfelvételre fordított idő megspórolásával, az eredeti nyelvű premierrel akár egy időben. A hangoskönyvek elhangozhatnak (elhunyt) híresség vagy szerzők hangján, sőt soha el nem mondott beszédek is hallhatóvá válnak az eredeti beszélő hangján (Martin é. n.). Ez utóbbit célozta a JFK Unsilenced projekt (W6), amelyben J. F. Kennedy megírt, de az 1963. november 23-án ellene elkövetett merénylet miatt el nem mondott dallasi beszédét hangosította meg a CereProc. A feladaton nyolc hétig dolgoztak: a közel 22 perces beszéd 831 elhangzott JFK-beszéd 116 777 fonetikai egységének elemzésével és felhasználásával készült el.

Televíziós és filmes produkciók is használ(hat)ják a technológiát: például a forgatás közben elhunyt szereplő karakterének pótlására (hasonló történt a CGI és a vizuális effektek segítségével az elhunyt Paul Walker esetén a *Halálos iramban 7* forgatása során – bár ez még nem kifejezetten az AI közreműködésével létrehozott módosítás volt), már nem élő színészek digitális feltámasztására vagy a veszélyesebb jelenetek kaszkadőr alkalmazása nélküli filmezésére (Hao 2020a).

A szórakozási élmény (hiper)perszonalizálhatóvá válik a deepfake révén: a felhasználó arca behelyettesíthető a főszereplő arca helyére, ezáltal még inkább im-

[4] A DALL-E elnevezés egyébként a Pixar animációs filmbeli robotjának, WALL-E-nek és Salvador Dalínak a nevéből született.

merzívvé válik a filmes és videójátékos élmény. (A kínai Zao app erre a lehetőségre épít; vö. Hao 2020a.) Ez a fajta személyre szabottság használható az e-kereskedelemben is (az online kereskedelem erősítését gazdaságilag fontos területnek tekintjük), hiszen a virtuális modellek felvehetik az adott vásárló külalakját, ily módon interaktív vásárlási élményt téve lehetővé, emellett az online vásárlás és a ruha virtuális felpróbálása során a testalkatra, személyes jegyekre is tekintettel tud lenni az internetes áruház (Hao 2020a; Lu 2019). Márpedig ha a fogyasztó a terméket saját tartozékaként látja, hajlandó többet venni belőle, magasabb árat fizetni érte, és másoknak is ajánlani a terméket. A sor hosszan folytatható: szinte hetente jelennek meg új és új alkalmazási módok, amelyek a mesterséges intelligencia segítségével történő kép- és hangszintetizálásra, illetve -manipulációra épülnek.

A Tencent nevű kínai technológiai vállalat egy jelentésében elismeri ugyan, hogy a deepfake képes károkat is okozni (például az abban nem szereplők arcának pornóvideókba történő montírozásával), ám a 2020-ban közzétett jelentés *(white-paper)* szerint a deepfake pozitívumait is mérlegre kell tenni. A technooptimista jelentés azt állítja, hogy a deepfake összességében nem lesz káros hatással a társadalomra, az igazságra, és „még kevésbé jelent veszélyt a világrendre" (Hao 2020a). Nem meglepő következtetés ez egy, a technológia kommercializálásából jelentősen profitáló vállalattól.

3.3. Alapján – nélkül

Az audiovizuális tartalom igazságértéke soha nem volt stabil, ráadásul a bizonyítékokat mindig is a társadalmi, politikai, kulturális viszonyok függvényében lehetett figyelembe venni (Paris–Donovan 2019). Az audiovizuális manipuláció körébe tartozik a legújabb technológiára építő, a mesterségesintelligencia-fejlesztés eredményeit hasznosító deepfake, de a Paris és Donovan (2019) fogalmával „cheap fake"-nek nevezett, már hagyományosnak számító kép-, hang- és videómanipulációs technikák sora is (gyorsítás, lassítás, kivágás, újrakontextualizálás stb.). Amire azonban mindkettő alkalmas: hatással tudnak lenni a vizuális bizonyítékok politikájára (*politics of evidence;* Paris–Donovan 2019). A(z audio)vizuális bizonyítékoknak mindig is kiemelt szerepük volt az újságírásban, a bírósági ítélkezésben, és az információhamisítás sem újdonság. A digitális kriminalisztika már jó ideje küzd a (digitális) tartalommanipuláció felderítésével. Amiben viszont újat hozott a deepfake és a mesterséges intelligencia használatával történő tartalommódosítás (sőt szövegtenyésztés, képgenerálás, hangszintetizálás), az a hamisítás tényének a viszonylag könnyű elfedése, a hamisítás technikailag egyszerű megvalósíthatósága, a detektálás körülményessége, illetve a digitális eredmények fénysebességgel történő terjedése. A bizonyítékok megkérdőjeleződése megingathatja a médiába, hatóságokba, bíróságba, sőt a tudományba vetett bizalmat is.

Kutatások (vö. Chesney–Citron 2018; Schiff et al. 2022) azonban már az ellenkező véglettel is foglalkoznak, mégpedig azzal, hogy az álhírekkel és deepfake-kel való riogatás egyes személyeknek még hasznos is lehet. A félretájékoztatásról szóló félretájékoztatás arra utal, hogy bizonyos személyek (különösen politikusok) számára előnyös a fake news és a deepfake jelensége, hiszen a hírnevüket negatívan érintő, ám valós információkat is álhír címkével tudják ellátni. Erre utal a magyar fordítással egyelőre nem rendelkező „Liar's dividend" 'a hazug jutaléka' fogalma (Chesney–Citron 2018), amely azt a paradoxont jelzi, hogy egyes politikusok profitálhatnak a félretájékoztatással telített információs környezetből (Schiff et al. 2022). Például egy, az adott személyt érő vádra a kommunikációs csapat manipulált videókat, hangfelvételeket készít és terjeszt, amelyekről természetesen kiderül, hogy hamisított anyagok. (Ez a közönségmanipulációs módszer kommunikációs szempontból a technológiai ellenőrizhetőség miatt nyilvánvalóan kockázatos.) Ezáltal a nagyközönségben kialakul a személyről készült videó- és hangfelvételek valódiságával kapcsolatban a kétely, emiatt tehát a személy az autentikus bizonyítékról is ki tudja jelenteni, hogy hamis. Ez pedig annál hihetőbb lesz, minél többet hall a nagyközönség a deepfake veszélyeiről (Chesney–Citron 2018). Ráadásul a politikus támogatói körében ez a „menekülőút" segíti a kognitív disszonancia elkerülését is, hiszen a negatív hírrel szemben is fenn tudják tartani a politikus támogatását (Taber–Lodge 2006).

Ebből adódik a dezinformáció nem direkt hatásaként a „kétely kételye" paradoxon, amely azt jelzi, hogy a felhasználók olyan kritikussá-cinikussá válnak a különböző forrásokkal kapcsolatban, hogy még a legitim információkat is megkérdőjelezik. Márpedig ha a laikus közönség megtanulja azt, hogy sem a szemének, sem a fülének nem hihet – még ha igaz is az információ –, az társadalmi-közéleti instabilitáshoz vezethet (Chesney–Citron 2018).

Ezt a folyamatot erősíthetik a 2023 elején gyors ütemben fejlesztett és piacra bocsátott AI-alapú csetrobotok, szövegtenyésztő alkalmazások, a meggyőzően gyors és a megadott stílust hihetően imitáló nyelvi modellek, amelyek azonban nem képesek a valóságot a hamis információtól elkülöníteni (Heaven 2022a), azaz sok esetben gyakorlatilag konfabulálnak: teljesen valótlan adatokat is valós információként, tényként állítanak be a generált szövegekben. Például a Meta által fejlesztett, 2022-ben a nagyközönség számára mindössze három napig elérhető Galactica nyelvi modell meggyőzően írt az űrben élő medvékről. Ez egy nagyon könnyen felderíthető és felfedezhető tévedés volt, ám olyan témákban, amelyekben valaki nem járatos, és az információknak nem jár utána, könnyedén megtévesztés áldozata lehet.

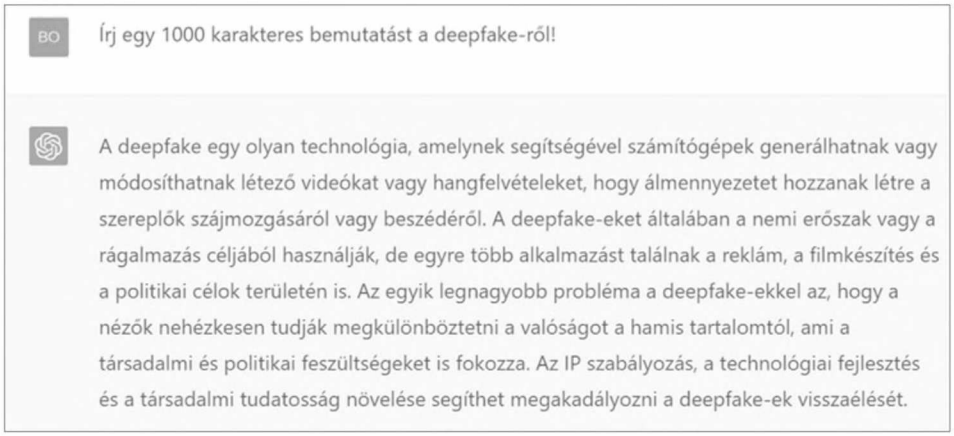
Sperber és munkatársai (2010) episztemikus éberségként *(epistemic vigilance)* nevezik meg azt a képességet, hogy a személy különbséget tud tenni megbízható és megbízhatatlan információ (hihető beszámoló és pletyka, bizalomra érdemes és hazug források) között. Az episztemikus éberség alapvetően két irányból működik: egyrészt a forrást (megbízhatóságát, hírnevét, esetleges rejtett célját), más-

részt a tartalmat (konzisztenciáját, más nézetekkel való koherenciáját) ellenőrzi. Ha működik az első mechanizmus, vagyis a hírek, nézetek, a videóban szereplő információk megfelelnek az ember intuícióinak, továbbá az elfogadottnak tekintett videók könnyebben hozzáférhető, külső jellegzetességeit is magukon hordozzák, megbízható forrásként tüntetik fel magukat, az csökkenti az éberséget.

Mindezek hatására a legnagyobb közösségimédia-oldalak, techcégek, hírszolgáltatók (és nyilvánvalóan a nemzetbiztonsági szervek is) próbálkoznak a megtévesztő tartalmak szoftveres felderítésével, címkézésével (és rosszindulatú szándék esetén azok letiltásával), ám az új és új manipulatív technológiák alapvetően az ezeket detektáló, leleplező programok előtt járnak.

4. ÖSSZEGZÉS

A témával foglalkozó, e fejezetben bemutatott tanulmányok többsége a deepfake okozta megtévesztés elleni védekezésben a médiatudatosság fontosságát, erősítését emeli ki – ám a kétely kételye paradoxon felhívja a figyelmet arra is, hogy a túlzott óvatosság, a minden információforrásban való kételkedés ugyanúgy veszélyeket hordoz magában.



3. ábra. A deepfake 1000 karakteres definíciója az AI-alapú csetalkalmazás, a ChatGPT szerint (forrás: saját képernyőfelvétel, 2023. január 30.)

Az AI generálta szövegek (különösen magyar nyelven) még tartalmaznak nyelvi vagy extralingvális árulkodó jeleket (vö. a deepfake ChatGPT-féle definícióját a 3. ábrán), ám a tanuló algoritmusok révén a tenyésztett szövegek minősége is folyamatosan fejlődik. Az egyre jobb minőségben előállított deepfake-videók az emberi szem számára megtévesztőek tudnak lenni – a hagyományos felismerési

módok (például: a tárgyak széle pixeles; a hang és kép nincs szinkronban; a beszélő vagy valamely testrésze mozdulatlan; nagyon kevés arcmozgás; a beszélő szájmozgása nem természetes; élettelen szemek; nincs vagy szokatlan pislogás) egyre kevésbé szolgálnak támpontul a mesterséges intelligencia segítségével manipulált tartalmak felismerésében. Nem túlzás tehát azt állítani, hogy az AI alkotta tartalmak valódiságát hamarosan AI vezérelte szoftverek fogják ellenőrizni. A mi feladatunk erre is felkészülni.

SZAKIRODALOM

Ajder, H. – Patrini, G. – Cavalli, F. – Cullen, L. 2019: The State of Deepfakes: Landscape, Threats, and Impact. *Deeptrace*. https://deeptracelabs.com/resources/ [2020. 04. 18.; már nem elérhető: 2023. 01. 30.]

Bennett, W. Lance – Livingston, Steven 2018. The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions. *European Journal of Communication* 33/2: 122–139.

Bicsérdi-Fülöp Ádám 2022: Egyedi műalkotássá alakítja a Dalí Múzeum AI-ja a látogatók álmait. *Kreatív*, november 29. https://kreativ.hu/cikk/egyedi-mualkotassa-alakitja-a-dali-muzeum-ai-ja-a-latogatok-almait [2023. 01. 30.]

Borsos, Zalan – Marinier, Raphael – Vincent, Damien – Kharitonov, Eugene – Pietquin, Olivier – Sharifi, Matt – Teboul, Olivier – Grangier, David – Tagliasacchi, Marco – Zeghidour, Neil 2022: AudioLM: a Language Modeling Approach to Audio Generation. Google Research. *arXiv* 2209.03143v1 [cs.SD], szeptember 7.

Bradshaw, Samantha – Howard, Philip N. 2018: The Global Organization of Social Media Disinformation Campaigns. *Journal for Internal Affairs* 71: 23–31.

Chesney, Robert – Citron, Danielle Keats 2019: Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security (July 14, 2018). 107 California Law Review 1753 (2019), U of Texas Law, Public Law Research Paper No. 692, U of Maryland Legal Studies Research Paper No. 2018-21. https://ssrn.com/abstract=3213954; http://dx.doi.org/10.2139/ssrn.3213954

Damiani, Jesse 2019: A Voice Deepfake Was Used To Scam A CEO Out Of $243,000. Forbes, 2019. szeptember 3. https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

Dobber, Tom – Metoui, Nadia – Trilling, Damian – Helberger, Natali – de Vreese, Claes 2020: Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes? *The International Journal of Press/Politics* 1–23. doi: 10.1177/1940161220944364

Feathers, T. 2020: Facial recognition company lied to school district about its racist tech. *Vice Motherboard*, december 2. https://www.vice.com/en/article/qjpkmx/fac-recognition-company-lied-to-school-district-about-its-racist-tech

Flynn, D. J. – Nyhan, Brendan – Reifler, Jason 2017: The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs about Politics. *Political Psychology* 38: 127–150.

Foer, Franklin 2018: The Era of Fake Video Begins. *The Atlantic*, május. https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/

Hao, Karen 2020a: The owner of WeChat thinks deepfakes could actually be good. *MIT Technology Review*, július 28. https://www.technologyreview.com/2020/07/28/1005692/china-tencent-wechat-ai-plan-says-deepfakes-good/

Hao, Karen 2020b: Memers are making deepfakes, and things are getting weird. *MIT Technology Review*, augusztus 28. https://www.technologyreview.com/2020/08/28/1007746/ai-deepfakes-memes/

Hao, Karen 2021: This is how we lost control of our faces. *MIT Technology Review*, február 5. https://www.technologyreview.com/2021/02/05/1017388/ai-deep-learning-facial-recognition-data-history/

Harwell, Drew 2020. Dating apps need women. Advertisers need diversity. AI companies offer a solution: Fake people. *Washington Post*, január 7. https://www.washingtonpost.com/technology/2020/01/07/dating-apps-need-women-advertisers-need-diversity-ai-companies-offer-solution-fake-people/ [2020. 10. 14.]

Heaven, Will Douglas 2021: How AI is reinventing what computers are. *MIT Technology Review*, október 22. https://www.technologyreview.com/2021/10/22/1037179/ai-reinventing-computers/

Heaven, Will Douglas 2022a: Why Meta's latest large language model survived only three days online. *MIT Technology Review*, november 18. https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/

Heaven, Will Douglas 2022b: The dark secret behind those cute AI-generated animal images. *MIT Technology Review*, május 25. https://www.technologyreview.com/2022/05/25/1052695/dark-secret-cute-ai-animal-images-dalle-openai-imagen-google/

Horváth Evelin – Mezriczky Marcell 2021: Új valóság születik. A CGI és a deepfake lehetőségei és veszélyei a médiatudatosság tükrében. *Magyaróra*, 1: 35–40.

Huffman, Mark 2020: Voice cloning could make the grandparent scam more dangerous. *Consumer Affairs*, február 25. https://www.consumeraffairs.com/news/voice-cloning-could-make-the-grandparent-scam-more-dangerous-022520.html

Johnson, Khari 2023: Iran Says Face Recognition Will ID Women Breaking Hijab Laws. *Wired*, január 10. https://www.wired.com/story/iran-says-face-recognition-will-id-women-breaking-hijab-laws/ [2023. 01. 12.]

Lee, Alex 2020: This ugly t-shirt makes you invisible to facial recognition tech. *Wired*, május 11. https://www.wired.co.uk/article/facial-recognition-t-shirt-block

Martin, Kim é. n.: What is Voice Cloning? https://www.idrnd.ai/what-is-voice-cloning/

Mezriczky Marcell 2021: *Ál/arc – A deepfake és a digitális videómanipuláció.* Mesterszakos szakdolgozat, kézirat. Budapest: Budapesti Corvinus Egyetem, Kommunikáció- és Médiatudomány Tanszék.

Nyiri Donát 2022: Átalakul a grafikusi szakma? Az AI art nem jön, hanem már itt van! *Minner*, október 3. https://minner.hu/atalakul-a-grafikusi-szakma-az-ai-art-nem-jon-hanem-mar-itt-van/

OECD 2019: *Artificial Intelligence in Society*. Paris: OECD Publishing.

Öhman, Carl 2020: Introducing the Pervert's Dilemma: A Contribution to the Critique of Deepfake Pornography. *Ethics and Information Technology*, 22: 133–140.

Özkiziltan, Didem – Hassel, Anke 2021: *Artificial Intelligence at Work: An Overview of Literature*. Governing Work in the Digital Age Project Working Paper Series.

Paris, Britt – Donovan, Joan 2019: Deepfakes and cheap fakes. The manipulation of audio and visual evidence. *Data & Society*, szeptember 18. https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1.pdf

Raji, Inioluwa Deborah – Fried, Genevieve 2021: About Face: A Survey of Facial Recognition Evaluation. *ArXiv* abs/2102.00813: n. o.

Rothman, Joshua 2018: In the Age of A.I., Is Seeing Still Believing? *The New Yorker*, november 5. https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing

Ryan-Mosley, Tate 2021: Beauty filters are changing the way young girls see themselves. *MIT Technology Review*, április 2. https://www.technologyreview.com/2021/04/02/1021635/beauty-filters-young-girls-augmented-reality-social-media/

Schiff, Kaylyn J. – Schiff, Daniel S. – Bueno, Natalia 2022: The Liar's Dividend: The Impact of Deepfakes and Fake News on Trust in Political Discourse. *OSF*, november 8. doi: 10.17605/OSF.IO/QPXR8.

Snow, Jackie 2017: AI Could Set Us Back 100 Years When It Comes to How We Consume News. *MIT Technology Review*, november 7. https://www.technologyreview.com/2017/11/07/147903/ai-could-send-us-back-100-years-when-it-comes-to-how-we-consume-news/

Snow, Olivia 2022: 'Magic Avatar' App Lensa Generated Nudes From My Childhood Photos. *Wired*, december 7. https://www.wired.com/story/lensa-artificial-intelligence-csem/

Sperber, Dan – Clement, Fabrice – Heintz, Christophe – Mascaro, Olivier – Mercier, Hugo – Origgi, Gloria – Wilson, D. 2010: Epistemic vigilance. *Mind & Language*, 25/4: 359–393.

Taber, Charles S. – Lodge, Milton 2006: Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50: 755–769. doi: 10.1111/j.1540-5907.2006.00214.x

Veszelszki Ágnes 2021: deepFAKEnews: Az információmanipuláció új módszerei. In: Balázs László (szerk.): *Digitális kommunikáció és tudatosság*. Budapest: Hungarovox Kiadó. 93–105.

Xu, Tammy 2022: Google's new AI can hear a snippet of song—and then keep on playing. *MIT Technology Review*, október 7. https://www.technologyreview.com/2022/10/07/1060897/ai-audio-generation/

FORRÁSOK

Grant, Kirsty 2021: Influencers react to Norway photo edit law: ‚Welcome honesty' or a 'shortcut'? *BBC*, 2021. 07. 06. https://www.bbc.com/news/newsbeat-57721080 [2021. 07. 15.]

Lee, Sze-Fung – Fung, Benjamin C. M. 2022: Think deepfakes don't fool you? Sorry, you're wrong. *The Next Web,* május 9. https://thenextweb.com/news/deepfakes-fool-you

Lu, Victor 2019: AI Creates Fashion Models With Custom Outfits and Poses. *Synced Review*, augusztus 29. https://medium.com/syncedreview/ai-creates-fashion-models-with-custom-outfits-and-poses-a27d5784651f

Mollick, Ethan (Emollick) 2022: Twitter, november 27. https://twitter.com/emollick/status/1596579802114592770

Sholihyn, Ilyas 2020: Someone deepfaked Singapore's politicians to lip-sync that Japanese meme song. *AsiaOne*, augusztus 7. https://www.asiaone.com/digital/someone-deepfaked-singapores-politicians-lip-sync-japanese-meme-song

Wilson, Mark 2019: The Salvador Dalí Museum just Deepfaked Dalí–see the video here. *Fast Company*, május 28. https://www.fastcompany.com/90355299/the-salvador-dali-museum-just-deepfaked-dali-see-the-video-here

W1 = Midjourney. https://www.midjourney.com/

W2 = Stable Diffusion. https://stability.ai/blog/stable-diffusion-public-release

W3 = Dall-e 2. https://openai.com/dall-e-2/

W4 = Imagen, Google. https://imagen.research.google/

W5 = This person does not exist. https://thispersondoesnotexist.com/

W6 = JFK Unsilenced. https://www.cereproc.com/en/jfkunsilenced

ACZÉL PETRA

A deepfake mint hazugság: együttműködés a megtévesztésben

A médiában megjelenő deepfake egy személy vagy esemény szimulált, valóságvonatkozással ugyan bíró, de saját realitás nélküli változata, és mint ilyen – megtévesztő. A mélytanuló programozás révén előállított audiovizuális tartalmak egyfajta hazugságként is értelmezhetők tehát, amelyeket a befogadó igen nehezen tud leleplezni. Ennek azonban jellemzően nem a megtévesztés tökéletes kivitelezése az oka, sokkal inkább az, hogy az emberi kommunikációt alapjaiban jellemzi a felek együttműködési szándéka. Ez az a szándék, amely lehetővé teszi, hogy a befogadó az interakcióhoz való hozzájárulásával voltaképpen maga is jóváhagyja a hazugságot. A fejezet a megtévesztés kommunikációelméleti vonatkozásait bemutatva azt tűzi ki célul, hogy rávilágítson a deepfake-tartalmak befogadásának, a megtévesztés elfogadásának hátterére. A deepfake üzenetközpontú definíciójából kiindulva előbb az információ-, majd az interakció-központú hazugságelméleteket veszi számba azért, hogy végül ezekkel írja le a deepfake-et mint hazug interakciót.

Kulcsszavak: hazugságelmélet, információmanipuláció, interakció, megtévesztés, együttműködés

1. PROLÓGUS

165 évig működött Amerika egyik legelismertebb galériája. New York szívében, a művészetkedvelő gazdagok elegáns „találkahelyeként" szolgáló Knoedler előbb régi műalkotásokkal, később aztán kortárs festőkkel foglalkozott. Többek között Diebenkornnal, Frank Stellával vagy Scullyval. A Knoedler nem csupán dollármilliomosoknak kínált festményeket, hanem múzeumoknak is, igen magas összegekért. Ma azonban hiába keresnénk Manhattanben a neves művészeti központot. 2011-ben ugyanis váratlanul bezárt. Mégpedig azért, mert a Knoedler nevéhez fűződik az ezredforduló legnagyobb festményhamisítási botránya. 1994 és 2008 között ugyanis a galéria – Ann Freedman elnökségével és irányításával – mintegy negyven hamis absztrakt expresszionista művet vett meg és adott el. Kivételesen nagy árréssel. Legtöbbjük Jackson Pollocknak és Mark Rothkónak tulajdonított alkotás volt. Több mint 80 millió dolláros csalásról volt szó, amelyben a galéria

vezetője ártatlannak vallja magát azóta is. Azokra a szakértőre hivatkozik, akiket felkért véleményezésre, miután ugyanattól a rejtélyes múltú hölgytől befogadta a festményeket, sok éven át. Valamennyi, általa megkérdezett hozzáértő egyöntetűen azt állította, hogy eredeti művekről van szó. Később más szakértők kiderítették, hogy mégsem. Még a legtapasztaltabbakat is megvezette egy tehetséges, kínai származású piktor, aki az amerikai nagyváros egyik garázsában alkotott – mások nevében. Értelemszerűen nem csak őket vezették meg, és nem csak a 2000-es évek elején. Mondják – legalábbis így tesz Thomas Hoving, a New York-i Metropolitan Múzeum egykori művészeti igazgatója, *False Impressions* című könyvében (Hoving 1997) –, hogy a Metropolitanben kiállított tárgyak mintegy 40 százaléka hamisítvány. Ezért is az ismert élcelődés, miszerint Camille Corot (1796–1875) összesen 3000 művet festett, amelyből 5000 az Egyesült Államokban található.

A festményhamisítás századok óta velünk élő jelenség, komoly művészettörténeti, szakmai és gazdasági kihívás, egyszersmind megszokott és hallgatólagosan feltételezett csalás. A jelenség ágense, a festményhamisító pedig igazi mélytanuló: az elérhető adatok sokaságának, a megismerhető módszereknek és a hozzáférhető, hiteles fizikai anyagoknak (festékek, vászon) és eszközöknek használatával hozza létre a duplikátot – vagy éppen a régi mesternek tulajdonított vadonatúj alkotást. A jó hamisító, ha sok-sok tanulmányozás és minden összegyűjthető információ alapján megtanulta az utánzott alkotót, akkor sokféle példányt tud létrehozni. Amiben mégis eltér korunk új jelenségétől, a deepfake-imágótól, az az, hogy minden egyes munkája egyedi, a maga módján megismételhetetlen. A festményhamisító egyedi csaló, a deepfake a tömeges módszer, előbbi alkotást, utóbbi eljárást hoz létre. Számonkérhető-e a csalás felismerése a deepfake-reprezentáción, ha a művészettörténésztől sem várható el teljes biztonsággal? Bele kell-e nyugodnunk, hogy a deepfake a mi jóváhagyásunkkal csal?

E fejezet célja, hogy hipotetikusan megvizsgálja, miért nehéz a mélycsalás eredményeiben, a képernyőn megjelenő szintetikus imágókban a megtévesztést, a hazugságot tetten érni. Abból a feltételezésből indul ki, hogy a megtévesztés sikerét és hatékonyságát a tartalommal-személlyel folyó interakció, a vélt együttműködési szándék befolyásolhatja, így a kommunikációban részt vevők mindegyike hozzájárulhat ahhoz, hogy a hazugság vagy torzítás ne derüljön ki. A tanulmány elsőként a deepfake fogalmát határozza meg röviden, majd taglalja a megtévesztéssel kapcsolatos kutatásokat, végül összeveti a deepfake-jelenséget a hazugságelméleti propozíciókkal.

2. A DEEPFAKE MINT TARTALOM

2022-ben Andrew Rossi rendezésében ismerhette meg a Netflix közönsége az *Andy Warhol Diaries* (Andy Warhol naplói) című dokumentumfilm-évadot. Ebben az ismert pop-art-művész maga olvassa fel naplóit. Warholról köztudott volt, hogy igen nehezen faggatható riportalany, híres volt az interjúiban adott rövid, egyszavas válaszairól. A film mégis azt sugallja, hogy másként is tudott a médiának beszélni. Sugallja, hiszen Warhol ugyan valóban egykor diktálta naplóit egy barátjának, de a hangot, amit a filmen hallunk, egy Resemble AI nevű cég dolgozta ki, csak a sorozat kedvéért. A már nem élő Andy Warhol hangját észleljük, azt, amelyen ő soha nem mondta akkor és azt, amit hallunk. Eközben az 1955-ben elhunyt James Dean negyedik (új) filmjét forgatja. Valójában természetesen nem ő, hanem a rendezőpáros, Anton Ernst és Tati Golykh, akik vele készítenek filmet. Egy, a vietnámi háború után szélnek eresztett több ezer katonai kutyáról szóló műhöz kerestek „másodlagos férfi főszereplőt". Hiába kutattak a ma élők között, végül – ahogyan ők fogalmaztak – a túlvilágról választották az egyetlen alkalmast, James Deant. A színészt családja engedélyével fogják posztumusz alkalmazni a feladatra.

Ezúttal tekintsünk el a fenti esetek technológiai, jogi, etikai vonatkozásaitól, inkább arra az aspektusra koncentráljunk, hogy mindkét alkalommal szintetikus, a személy létét halálában állító, megtévesztésre alkalmas tartalom jött létre, olyan üzenet, amely hazugságnak is tekinthető.

A deepfake a mélytanulással létrehozott hazug tartalom jellemzője. Azért hazug, illetve megtévesztő, mert nem új – addig nem létező – virtuális személyt vagy jelenséget alkot meg, hanem egy meglévőt utánoz, szimulál. Utal ugyan egy referenciális létezőre, hiszen újragenerálja a személyt vagy tárgyat, de nincs saját realitása, nem hiteles (vö. Baudrillard 1996). Meghatározásai közül tehát a tanulmány céljaira azokat említem, amelyek a deepfake-re mint üzenetre, mint közvetítőre (médiára) fókuszálnak. Ezek szerint a deepfake olyan szintetikus – értsd művi módon előállított, a valóságban vagy természetben nem létező – média, amelyet a mesterséges intelligencia generál, amely alkalmas olyan események és jelenségek ábrázolására, amelyek nem történtek meg, de alkalmasak egy személy hitelességének rombolására (is) (Schick 2020; Whittaker 2020; Kalpokas 2021). A reprezentáció tekintetében egy személy hipervalóságos digitális másolata, amely minden tekintetben manipulálható (Hughes et al. 2021). Másképpen fogalmazva, a deepfake a szintetikus média egy speciális fajtája, amelyben egy képen vagy videón szereplő személyt egy másik személy képmásával cserélnek ki (Somers 2020). Összefoglalóan a deepfake „gépi és mélytanulással dolgozó AI által létrehozott manipulált vagy szintetikus audiális-vizuális média, amely hitelesnek tűnik, és amelyben a megjelenő személyek azt mondják vagy teszik, amit a valóságban nem mondtak/tettek" (European Parliamentary Research Service 2021: 1).

3. HAZUGSÁG ÉS MEGTÉVESZTÉS A KOMMUNIKÁCIÓBAN

Az emberi kommunikációs folyamatok egyike sem írható le tisztán hazugságként vagy igazságközlésként, a kommunikátorok mindegyike valamilyen mértékben manipulálja a közlésfolyamatban megjelenő információt, amely ez által részben őszintévé, részben megtévesztővé válhat (McCornack 1992). A hazugság az emberi interakciók sajátossága, a közlésfolyamatok egyharmadában megjelenik. Ez a számosság kommunikátoronként napi két hazugságot feltételez (DePaulo et al. 1996). Az elmúlt évtizedek kutatásai pedig rendre azt állapították meg, hogy az ember hazugságfelismerő képessége általában gyenge, leginkább azért, mert megszokott döntési mintákra vagy intuícióra és kiválasztott jelek megfigyelésére alapozzák megítélésüket a közlővel kapcsolatban (Verschuere et al. 2023).

A megtévesztésnek számos módja létezik (Metts 1989; Turner–Edgley–Olmstead 1975; Veszelszki 2021), ezek között szerepel a valótlan információ állítása, csakúgy, mint a homályosítás, a lényegtelen információ kiemelése, az információ egy részének elhagyása. Herbert Paul Grice (1989) a kommunikáció során a felek részéről valamilyen mértékben kölcsönösen megnyilvánuló szándék kapcsán rögzítette az együttműködés elvét, amely megköveteli az ún. társalgási maximák betartását. Négy ilyen társalgási alapszabályt azonosított. A mennyiség szabálya arra vonatkozik, hogy a kommunikáció informatív legyen, pontosan annyit közöljön, amennyi szükséges. Ha azt kérdezzük meg valakitől, hogy találkozott-e egy közös ismerősünkkel, akkor az informativitás jelenthet egy egyszerű igenlést vagy tagadást. Ugyanakkor, ha később kiderülne, hogy ez az ismerős a találkozás során nekünk is üzent valamit, amit a válaszadó elfelejtett továbbadni, akkor akár elhallgatásnak is tekinthetjük tömör igenlő válaszát. Másodikként a minőség maximája a közlés igazságtartalmát írja elő: azt, hogy olyan információt ne közöljünk, amelyről tudjuk, hogy hamis, vagy amelyre nincs bizonyítékunk. Ahogyan az előző példában az együttműködés alapelvéből kiindulva azt feltételezhetjük, hogy beszélgetőpartnerünk minden szükséges közléstartalmat megosztott velünk, úgy a minőség maximája alapján azt gondolhatjuk, hogy nem fogja azt mondani találkozás esetén, hogy mégsem találkozott közös ismerősünkkel. A viszony maximája azt diktálja, hogy a felek olyan információt osszanak meg egymással, amely a megelőző közlésekhez lényegileg, relevánsan kapcsolódik. Választ adnak, ha kérdezték őket, olyan kijelentéseket tesznek, amelyek kapcsolódnak korábbi kijelentéseikhez. A modor maximája szerint pedig a közlésfolyamat résztvevői kerülni fogják az érthetetlenséget, a homályosságot, a kétértelműséget, ezzel a kommunikáció módjára ügyelve. A hétköznapi személyközi és nyilvános kommunikációs helyzetekben e szabályok együttese kevéssé teljesül. Ugyanakkor a kommunikáló felek, az együttműködés szándékát feltételezve afelé orientálódnak, hogy a szabályok követését feltételezzék beszédpartnereiktől (Levinson 1983).

Az információmanipuláció elmélete szerint a megtévesztő üzenetek a társalgási maximák rejtett megsértésének során jönnek létre (McCornack 1992). Ezek esetében a befogadó fél azt vélelmezi, hogy a küldő betartja az együttműködési alapelvet, és követi a maximákat. A küldő, vagyis a megtévesztő fél pedig tartja magát ehhez az észleléshez, miközben, rejtve, vét a maximák ellen (Galasiński 2000). Mindezek alapján legalább négy módon lehetséges megtévesztően kommunikálni: 1. a közölt információ mennyiségével, az informativitás mértékével, 2. a valótlan információ közlésével, 3. az üzenet relevanciájának, más tartalomhoz való viszonyának manipulálásával és 4. az információ bemutatásának stílusával, a mód alakításával. A hazugság során a befogadót egyrészt saját feltételezése téveszti meg, mely szerint az üzenet követi az együttműködési alapelveket: informatív, igaz, releváns és érthető (McCornack et al. 1996). Másrészt a befogadót az is megtéveszti, hogy miközben a hazugságot hallja, látja, olvassa, egyidejűleg olyan további információk meglétét feltételezi, amelyek nem igazak. Ha a fenti példánál maradunk: amennyiben partnerünk kérdésünkre csupán azt válaszolja, hogy találkozott közös ismerősünkkel, akkor ezt hallva azt feltételezzük, hogy semmi olyasmi nem történt, ami minket ennél közelebbről érintene. Ez a feltételezés azonban nem helytálló, hiszen a közös ismerős üzenetet küldött nekünk, amit partnerünk válasza nem tartalmazott. Ez esetben két tekintetben is megtévesztés történik: egyfelől azért, mert a befogadó az együttműködési alapelv maximáinak teljesülését vélelmezi, másfelől, mert további, nem helytálló információkat feltételez az esettel kapcsolatban.

A hazugság tudományos feldolgozásában máig meghatározóak a pszichológiai megközelítések, így többek között Ekmannak és Friesennek (1969) a nem nyelvi szivárogtatásról szóló elmélete. Ez abból a feltételezésből indul ki, hogy a testi kommunikáció kevésbé képes elrejteni a hazugságot, akár mások, akár önmagunk megtévesztéséről van szó. Az arcon ilyen árulkodó jelek lehetnek például az üzenettől eltérő, önkéntelenül vagy már tudatosan, de még tökéletlenül kifejezett érzelmek, a túlzó, hosszan kitartott faciális reakciók, a kezek, karok esetében az arcjátékkal összhangban nem álló mozdulatok. Hasonlóképpen ismert Zuckerman, DePaulo és Rosenthal (1981) négyfaktoros elmélete, amely szerint a hazugságot négy olyan érzelmi-mentális mechanizmus kíséri, amelyek árulkodó jelek forrásaiként is szolgálhatnak, így az izgalom (például, mert a megtévesztés a közlő saját értékszemléletével konfliktusban áll, vagy mert félelmet érez, hogy a hazugságra fény derül), a viselkedés kontrolligénye (ez is szivárgáshoz vezethet, például amikor az arcmozgások nagyon ellenőrzöttek, a lábmozgások viszont nem), a hazugsággal kapcsolatos érzelmek (ilyen lehet például a megtévesztés sikere felett érzett örömből fakadó önkéntelen és az interakcióban nem relevánsan megjelenő mosoly), a kognitív erőfeszítés (a hazugság ugyanis nagyobb koncentrációt kíván a rejtés miatt, ami hosszabb fogalmazáshoz, több szünet alkalmazásához vezethet). Ezek az elméletek ugyanakkor a hazug közlő belső pszichológiai folyamatai-

ra, illetve külsőleg érzékelhető, leginkább szándéktalan megnyilvánulásaira utalnak. A kommunikációs folyamatban létrejövő megtévesztést ugyanakkor számos szándékos, kontextuális jegy is befolyásolja. Ezekre mutat rá a személyközi megtévesztés elmélete (Buller–Burgoon 1996; Buller–Burgoon 2004). Az elmélet egyik kiindulópontja, hogy az emberi kommunikációnak, beleértve a megtévesztést is, számtalan célja van, úgymint a kedvező énbemutatás, a harmonikus kapcsolatfenntartás, a konverzáció megkönnyítése, a meggyőzés vagy az azonosulás lehetősége. Az interakcióban a felek tehát adnak és kapnak, és ez alapvetően szándékoltan zajlik így. A másik kiinduló állítás, hogy az információmenedzsment alapvető a kommunikációban. Vagyis az emberek elrejtenek, torzítanak, homályosabbá tesznek vagy éppen túlhangsúlyoznak információrészleteket; ezzel a közlés megbízhatóságát, nyíltságát, relevanciáját, személyességi fokát alakítva. A harmadik alapvetés, hogy a befogadók nem passzív, hanem aktív résztvevői, alakítói a megtévesztő kommunikációnak. Bizalomszintjüknek vagy éppen gyanakvásuk mértékének megfelelően vesznek részt az ilyen interakciókban, és járulnak hozzá azok kimenetéhez.

Buller és Burgoon huszonegy igazolható propozíciót dolgozott ki a felek együttműködésével létrejövő megtévesztés vizsgálatára és leírására. Ezúttal ezekből a témánkhoz relevánsan kapcsolódókat mutatom be részletesebben, a többit csupán érintőlegesen. Az első propozíció lefekteti, hogy a médium interaktivitása, illetve a társalgásra vonatkozó igény adják a megtévesztő üzenetváltások kontextusának azon jellemzőit, amelyek szisztematikusan befolyásolják a küldő és a fogadó kogníciót és viselkedését. E szerint lényegi szerepe lehet az interaktivitásnak, vagyis annak, hogy a kapott üzenet kapcsolódik-e korábbi üzenetekhez, illetve a médium mennyire engedi az egyidejűséget és a nyelvi, illetve nem nyelvi jelek közvetítését. Szintén fontos tényező lesz, hogy a társalgás mentálisan és emocionálisan menynyire fenntartható. A második propozíció pedig azt foglalja magában, hogy a felek közötti kapcsolat tekintetében az ismerősség és a kapcsolat értékessége lesz a küldő és befogadó kogníciójá és viselkedését befolyásoló tényező. Minél több kommunikációs jelhez férünk hozzá a médiumon keresztül, annál sikeresebb lehet a küldő részéről a hazugság. Minél korlátozottabb a médium jelátvivő hatékonysága, annál több lehet az árulkodó jel (például aközött, hogy a küldő szövegesen leírja vagy videóüzenetben elküldi megtévesztő közlését, a második sikere valószínűbb). Egyúttal, ha a befogadó jól ismeri a küldőt, nem vezet jellemzően ahhoz, hogy hatékonyabban venné észre a hazugságot. Ugyanis a küldő, éppen az ismerősség miatt, jóval pontosabban és valószínűbben tudja kidolgozni megtévesztő üzenetét. A harmadik propozíció szerint az interaktív kontextusokban és a pozitívan hangolt kapcsolatokban azt várjuk, hogy a küldő igazat mond, míg a negyedik propozíció arra mutat rá, hogy annál kevésbé fél a megtévesztő a lelepleződéstől, minél inkább tudja, hogy a befogadó ismeri őt és a megtévesztő viselkedést, és arra számít részéről, hogy igazat mond. Az 5., 6., 7., 8. propozíciók

a megtévesztésben megjelenő stratégiai és nem stratégiai tevékenységeket taglalják. Az előbbiek leginkább az információmanipuláció módjaira vonatkoznak, az utóbbiak olyan árulkodó jelekre, mint az izgalom vagy a félelem. A kilencedik propozíció azt fejti ki, hogy a felek közötti közelebbi ismerősség a küldőkből több stratégiai és nem stratégiai kommunikációs cselekvést vált ki, a tizedik pedig arra utal, hogy a felkészült, kompetens küldő felek kommunikációja több stratégiai és kevesebb nem stratégiai elemet tartalmaz, mint a felkészületleneké. A 11.-től a 14. propozícióig az alapfeltevések négy faktorra hívják fel a figyelmet. Arra, hogy a befogadók nagyobb valószínűséggel tekintik a küldőt hitelesnek, egyrészt, ha a kontextus-csatorna interaktív, másrészt, ha a befogadó elfogult a küldő hitelességével kapcsolatban, és harmadrészt amennyiben a küldő gyakorlott és felkészült kommunikátor. A negyedik faktor a küldőtől megszokott kommunikációs mintákra vonatkozik. Ha ettől a küldő eltér, illetve ha a befogadó ismeri a küldő által felhasznált információk egy részét és a küldő viselkedését, illetve ha a befogadónak jók a dekódolási, értelmezési képességei, akkor nagyobb valószínűséggel lesz képes felfigyelni a megtévesztésre. A következő négy propozíció (15–17.) azt taglalja, hogy a befogadóban mikor keletkezik gyanú a küldő hitelességével kapcsolatban: leginkább akkor, ha a közlés eltér a küldőtől megszokott eddigi viselkedéstől, ha a küldő bizonytalanságot sugall, illetve ha az üzenet információmennyisége nem megfelelő a befogadó számára. Amint a gyanakvás megjelenik, az a stratégiai és nem stratégiai cselekvésre egyaránt hatással lesz. Az utolsó négy (18–21.) propozíció a kommunikációs folyamatra vonatkozik, arra, hogy a küldő hitelességének végső megítélése és a befogadó hazugságfeltárásának pontossága a korábban már említett elfogultságtól, gyanakvástól, a küldő záró viselkedésmozzanataitól is függ. Az elméletalkotók az elmúlt tizenhét esztendőben számos alkalommal tesztelték a fenti propozíciókat és igazolták helytállóságukat (többek között White–Burgoon 2001; Burgoon–Buller–Floyd 2002; Zhou–Burgoon–Twitchell–Nunamaker 2004).

4. A DEEPFAKE MINT HAZUGSÁG

A deepfake a személy plágiumaként is értelmezhető. A latin szó voltaképpen „emberrablást" jelent, a *plagiare* 'tőrbe ejt' igéből származik. Az eredeti értelem tehát pontosan írja le a mai jelenséget: a deepfake tőrbe ejti, elrabolja a – többnyire ismert – személyiséget. A deepfake tehát egyrészt a közlő személyét, másrészt a közlés tartalmát tekintve lehet megtévesztő. A grice-i együttműködési alapelv maximáinak teljesülése, illetve az erre épülő, McCornack által számba vett információmanipulációs módok a deepfake-üzenetben a személyre is vonatkoznak. Ilyen értelemben tehát a szimulált személy és mondanivalója együttesen tekinthető a deepfake-interakció tartalmának. A deepfake mint üzenet pedig vét a

maximák ellen, és alkalmazza az információmanipuláció módszereit. Nem tartja be az informativitás kívánalmát, mert többet közöl a szükségesnél, hiszen a valóság alternatíváját, mintegy többletét hozza létre, mind a valós személy szintetikus megkettőzésével, mind a közlendő tekintetében. Nem igaz, hiszen a közlő maga virtuális hasonmás, digitális utánzat, és nem releváns, mert az elhangzó szöveg nem kapcsolódik a közlőhöz vagy a valóságos közlő által korábban mondottakhoz. A deepfake-ben ugyanakkor a kommunikáció nem vagy nem mindig homályos, a mondanivaló általában érthető, a deepfake nem vagy kevéssé vét a modor maximája ellen. Amennyiben a híres politikai deepfake-eket vesszük – például Richard Nixon tartalékbeszédét arra az esetre, ha az 1969-es holdra szállás nem sikerülne (Moondisaster 2020), vagy Volodimir Zelenszkij megadásra felszólító videóját 2022 márciusából (Telegraph 2022) –, akkor látjuk, hogy az információmanipuláció az üzenet érthetőségét kivéve, mind a személyt, mind a mondanivalót illetően alkalmazza a relevancia, az informativitás és az igazság torzítását.

A deepfake-imágók sajátossága, hogy belső, „algoritmikus indentitásuk" – az elérhető és tanulható, valamint programozható információ mennyiségétől függően – is kiszivároghat a hazugság során. A deepfake is elárulja tehát önmagát, amennyiben a száj mozgása nem követi a hangzást, ha az arc mikromozgásai, a kéz mozdulatai eltérnek a személy addig ismert nonverbális viselkedésétől. A deepfake mint hazugság azonosítása mégsem gyakori, sem az észrevétel, sem a tudatosság nem küszöböli ki, hogy hasson a befogadóra (Hughes et al. 2021). Ennek okát a személyközi megtévesztés elméletének propozícióival magyarázhatjuk leginkább. Tekintettel arra, hogy a deepfake-imágók interaktív felületeken (közösségi médiában, videomegosztókon) érhetők el, olyan platformokon, ahol a befogadó is aktív résztvevő, illetve arra, hogy ezek a tartalmak többnyire ismert embereket plagizálnak, a hazugság befogadása, hatásának érvényesülése kevésbé ütközik akadályba (lásd az első két propozíciót). A mozgókép maga is élénk információtípus (Hill 2004), és ez intenzívebbé teszi a befogadói élményt, hasonlóan az ismerősséghez, amely a plagizált személy hírnevére épül. Mivel ismerjük a közlőt, ezért a hazugság azonosítása nehezebb, különösen akkor, ha az ismerősség pozitívan hangolt, amint azt a harmadik propozíció kimondja. A következő négy propozíció, amely a stratégiai és nem stratégiai cselekvéseket tárgyalja, a deepfake-re annyiban érvényes, hogy esetében a nem stratégiai – kiszivárogtató – jelek jóval kontrolláltabbak, elkerülhetők. Így, megfelelő kivitelezés mellett, a befogadó együttműködését nem akadályozzák gyanús jelek. A hazugságelmélet kilencedik propozíciója a kommunikációs felkészültségre hívja fel a figyelmet. A deepfake-imágókban, -hangokban minden üzenet előre, pontosan eltervezett, spontaneitása, így árulkodó hibája, tévesztése nincs. Ez kevésbé teszi lehetővé a befogadó számára, hogy a hazugságot észrevegye. A befogadó akkor is kész a megtévesztés elfogadására, ha a közlő hitelessége elfogulttá teszi, például egy presztízspozícióban lévő személy (elnök, miniszter, festő, tudós) esetében, de a hétköznapi emberek deepfake-jénél

is működhet ez a hatás. A lelepleződést a deepfake-imágó tökéletlensége segítheti; ha az ábrázolt személy viselkedése eltér a korábban megszokott mintáktól, ha bizonytalanságot sugall – ami a digitális reprezentáció hibáival magyarázható. Összegzésképpen, a deepfake-imagó hazugságának viszonylagos felismerhetetlenségét az interakció során kialakuló együttműködés támogatja. Minél inkább kapcsolatban érezzük magunkat a szimulált személlyel, minél inkább hiszünk az eredeti személyiségből vagy korábbi ismereteinkből fakadó hitelességben, minél nagyobb kapacitású médiumon találkozunk az üzenettel, annál valószínűbb, hogy nem fogjuk felismerni a mélycsalást. A kommunikálás maga akadályozhatja az igazság kiderülését. Ebben az értelemben tehát a deepfake megtévesztése a kommunikáló felek kooperációs hajlandóságát aknázza ki. Ez pedig a deepfake etikai megközelítésének egyik legfontosabb csomópontját adhatja.

5. EPILÓGUS

A festményhamisító ugyan alkot, a deepfake-programozó pedig gyárt, két vonatkozásban mégis hasonlítanak: szándékosan egy másik személyt másolnak, és megtévesztenek, hogy céljaik teljesüljenek. Különös módon azonban, mindkettejük esetében a befogadó együttműködik, vagyis részt vesz a hazugság létrejöttében, bizonyos tekintetben jóváhagyja, sőt értékeli a megtévesztést. Teszi ezt a személy hitelessége, a kapcsolat értéke, a csatorna interaktivitása miatt. Teszi ezt az ismerősség, a pozitív elvárások, a közlő tudatos, stratégiai cselekvése miatt, azon a természetes bizalom okán, amely nélkül az emberi interakció nehezen tartható fenn. Mindenesetre attól, hogy teszi, hozzájárul ahhoz, ami vele történik meg: a megtévesztéshez. Ennek a felelősségnek a felismerése fontos célkitűzése lehet a következő évek médiatudatosságra nevelésének, annál is inkább, mert az online bizalom (Etienne 2021) kérdései egyre égetőbbek mind a tudomány, mind a szabályozás számára.

SZAKIRODALOM

Baudrillard, Jean 1994: *Simulacra and Simulation*. Ann Arbor University of Michigan Press: Michigan.

Buller, David B. – Burgoon, Judee K. 1996: Interpersonal deception theory. *Communication Theory*, 6: 203–242.

Burgoon, Judee K. – Buller, David B. 2004: Interpersonal deception theory. In: Seiter, John S. – Gass, Robert H. (szerk.): *Perspectives on persuasion, social influence, and compliance gaining*. Boston, MA: Allyn & Bacon. 239–264.

DePaulo, Bella M. – Ansfield, Matthew E. – Bell, Kathy L. 1996: Theories about deception and paradigms for studying it: A critical appraisal of Buller and Burgoon's interpersonal decetion theory and research. *Communication Theory*, 6: 287–296.

Ekman, Paul – Friesen, Wallace. Non-Verbal Leakage and Clues to Deception. *Psychiatry. Journal for the Study of Interpersonal Processes*, 32/1: 88–106.

European Parliamentary Research Service 2021: *Tackling deepfakes in European policy*. Online: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf [2023. május 4.]

Etienne, Hubert 2021: The future of online trust (and why Deepfake is advancing it). *AI and Ethics*, 1: 553–562.

Galasiński, Darius 2000: *The Language of Deception. A Discourse Analyitical Study*. Thousand Oaks, CA: Sage.

Grice, H. Paul 1989: *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hill, Charles A. 2004: The Psychology of Rhetorical Images. In: Hill, Charles A. – Helmers, Marguerite (szerk.): *Defining Visual Rhetorics*. Mahwah, NJ: Lawrence Erlbaum Associates. 25–40.

Hoving, Thomas 1997: *False Impressions: The Hunt for Big-Time Art Fakes*. New York: Simon & Schuster.

Hughes, Sean – Fried, Ohad – Ferguson, Melissa – Hughes, Ciaran – Hughes, Rian – Yao, Xinwei – Hussey, Ian 2021: Deepfaked online content is highly effective in manipulating people's attitudes and intentions. PsyArXiv online. https://psyarxiv.com/4ms5a/ [2023. május 4.]

Kalpokas, Ignas 2021: Problematising reality: the promises and perils of synthetic media. *SN Social Sciences*, 1/1. https://link.springer.com/article/10.1007/s43545-020-00010-8

Levinson, Stephen C. 1983: *Pragmatics*. Cambridge MA: Cambridge University Press.

McCornack, Steven – Levine, Timothy R. –Morrison, Kelly – Lapinski, Mary 1996: Speaking of information manipulation: A critical rejoinder. *Communication Monographs*, 63: 83–91.

McCornack, Steven 1992: Information manipulation theory. *Communication Monographs*, 59: 1–16.

Metts, S. 1989: An exploratory investigation of deception in close relationships. *Journal of Social and Personal Relationships*, 6: 159–179.

Moondisaster 2020: In the event of moondisaster project. https://moondisaster.org/ [2023. május 4.]

Schick, Nina 2020: *Deepfakes: The Coming Infocalypse*. Twelve: New York.

Somers, Meredith 2020: Deepfakes, Explained. *Ideas Made to Matter, MIT.* https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained [2023. május 4.]

Telegraph 2022. https://www.youtube.com/watch?v=X17yrEV5sl4 [2023. május 4.]

Turner, Ronny E. – Edgley, Charles – Olmstead, Glen 1975: Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology*, 11: 69–89.

Verschuere, Bruno – Lin, Chu-Chien – Huismann, Sara – Kleinberg, Bennett – Willemse, Marleen – Jia Mei, Emily C. – Goor, Thierry van – Löwy, Leonie H. S. – Appiah, Obed K. – Meijer, Ewout 2023: The use-the-best heuristic facilitates deception detection. *Nature Human Behaviour*, March 2023. https://www.nature.com/articles/s41562-023-01556-2#citeas [2023. május 4.]

Veszelszki Ágnes 2021: deepFAKEnews: Az információmanipuláció új módszerei. In: Balázs László (szerk.): *Digitális kommunikáció és tudatosság*. Budapest: Hungarovox Kiadó. 93–105.

White, Cindy H. – Burgoon, Judee K. 2001: Adaptation and communicative design: Patterns of interaction in truthful and deceptive conversations. *Human Communication Research* 27: 9–37.

Whittaker, Lucas – Letheren, Kate – Mulcahy, Rory 2021: The Rise of Deepfakes: A conceptual framework and research agenda for marketing. *Australasian Marketing Journal*, 29/3: 204–214.

Zuckerman, Miron – DePaulo, Bella M. – Rosenthal, Robert 1981: Verbal and nonverbal communication of deception. In: Berkowitz, Leonard (szerk.): *Advances in experimental social psychology.* New York: Academic Press. 1–59.

MEZRICZKY MARCELL

Ne higgy a szemének!

A deepfake online sajtóreprezentációja 2018 és 2022 között

Az emberek hitelesebbnek tekintik a mesterséges intelligencia által generált arcokat hús-vér társaik megjelenésénél? A manipuláció és a valóság megkülönböztetéséhez a média által megalkotott konceptuális keretek, narratívák is hozzájárulnak, emiatt pedig fontos, hogy a deepfake-et milyen kontextusba ágyazzák. Kutatásom célja bemutatni, hogy az online sajtó milyen témákkal kapcsolja össze a technológiát, és hogyan változik annak reprezentációja az első magyar médiamegjelenéstől, 2018-tól a jelen vizsgálat idejére, 2022-re. A tartalomelemzés módszerével 882 db publikációt elemeztem, és ezek alapján elmondható, hogy számuk évről évre emelkedik, 2018 vizsgált időszakában még csak 25 db, 2022-ben pedig már 324 db találat született. A legtöbbször előforduló kategóriák a „kiberbiztonság", a „szórakoztatás" és a „bűncselekmény, fake news". Ennek jelentősége, hogy ha a pozitív felhasználási módok csekély mértékben szerepelnek a médiában, ha a veszélynarratívával találkozunk döntő többségben, akkor az hatással lehet arra is, hogyan fogadjuk be azokat a produktumokat, amelyek a lehetőséget láttatják a deepfake-ben.

Kulcsszavak: mélytanulás, mesterséges intelligencia, sajtóreprezentáció, tartalomelemzés

1. BEVEZETÉS

2023-ban talán nem túlzás kijelenteni, hogy a fizikai valóság vesztésre áll, a még főként elképzelés szintjén mozgó metaverzum nevű virtuális világ, az életünk részévé váló, számos, szépítő- és fizikai megjelenésen változtatni képes applikáció, továbbá a mesterségesintelligencia-alapú videómanipuláció, a deepfake következményeképpen, amely 2017 óta szerves része az életünknek. Gyakran azt mondjuk ezekre: szimulációk. Szimulációk, ugyanis a felsorolt elemek befogadása során elhitetik velünk, mintha lenne valamink, amink nincs (Baudrillard 2017) – például egy méregdrágán vásárolt telkünk a metaverzumban (Előd 2022). Ennek a jelentőségét egy, a színlelés és a szimuláció közti különbséget bemutató Baud-

rillard-analógia (2017) mentén lehet a legérzékletesebben kifejteni: ha betegséget színlelünk, azzal a realitás nem változik, nem vagyunk betegek, csupán eljátsszuk. Azonban, ha szimulálunk, képesek vagyunk néhány tünetet is produkálni, ez pedig „megkérdőjelezi az »igaz« és »hamis«, »valódi« és »képzeletbeli« közötti különbséget" (Baudrillard 2017). Így valamiféle valóságok közti realitás képződik: a szimulánsra nem tekinthetünk betegként, hiszen tulajdonképpen a tünetei nem ebből fakadnak, viszont azt sem mondhatjuk, hogy nem beteg, ugyanis tünetei vannak.

A deepfake is hasonló módon rekonstruálja a valóságot: tekinthetjük-e a Tom Cruise-t ábrázoló TikTok-videókat valódinak? Hiszen köztudott, hogy a technológiát felhasználva alkották meg őket, viszont a színész arcának kivételével minden részlet fizikai formában is megfelelt a valóságnak a felvétel pillanatában. A problémát az jelenti, ami Nightingale és Farid (2022) kutatásából is kiderült: az emberek hitelesebbnek tekintik a mesterséges intelligencia (MI) által generált arcokat hús-vér társaik megjelenésénél. A kísérletben részt vevők vegyes adathalmazon 48%-os pontossággal tudták eltalálni, melyik személy valódi, melyik nem; a szintetikus és a valós ábrázat láttán pedig előbbi 8%-kal magasabb bizalmi indexet ért el.

Abba a szakaszba lépett az arccserélős videók gyártása, amelyben már különösen nehéz megítélni, vajon algoritmus gyártotta produktumok vagy a fizikai valóságnak minden tekintetben megfelelő leképezések jönnek-e velünk szembe az interneten. Éppen ezért jutnak egyre hangsúlyosabb szerephez a deepfake különböző felhasználási módjai és az ezekről beszámoló médiamegjelenések. Egy ilyen magasfokú hitelességet imitálni képes technológia társadalmi megítélése nagyban függhet attól, hogy a média milyen narratívát kínál értelmezési keretként a fogyasztók számára. A magyar online sajtóban való reprezentáltság legjobb tudomásom szerint még nem képezte korábbi kutatás tárgyát a területen, így a tartalomelemzés módszerével arra vállalkoztam, hogy átfogó képet adjak a deepfake-ről szóló cikkek mennyiségi és minőségi jellemzőiről.

2. ELMÉLETI ÁTTEKINTÉS

A médiával és annak hatásaival kapcsolatban számos elméletet és modellt dolgoztak (és dolgoznak) ki. Egyes gondolkodók passzív félként (lásd Harold Lasswell és a „lövedékelmélete", 1927), mások aktív cselekvőként (például Blumer és Katz „használat és kielégülés" modellje, 1974) tekintenek a befogadóra, a média és fogyasztója kapcsolatának kontextusában (Aczél–Andok–Bokor 2015). Afelől azonban nincs kétség, hogy a média jelentős hatást gyakorol mindennapi életünkre a társadalmi érintkezések minden szintjén. A jelen fejezet alapjául is szolgáló keretezést (framinget) két, egymással kapcsolatban álló, mégis különböző perspektívából mutatom be.

2.1. A keretezés értelmezései

Bajomi-Lázár Péter 2006-ban gyűjtötte össze a médiahatás-kutatás legfontosabb mérföldköveit, elméleteit *Manipulál-e a média?* című cikkében. A szerző a keretezés elméletét Chomsky és Herman nevéhez köti; és elsősorban médiahatás-modellként jeleníti meg. A megközelítés röviden összefoglalható: a média (amely az elitek kezében van, a hétköznapi ember csak passzív befogadó) úgy befolyásol minket a hírek által, hogy az egyes csatornákon bemutatott információkat konceptuális „keretben" elhelyezve tárja elénk (Bajomi-Lázár 2006). Tehát nem az információ „tiszta" változatát adják a befogadó birtokába, hanem annak egy adott kontextusban elhelyezett, módosított verzióját. A közönség ezáltal könnyebben dolgozza fel az adatokat (és olyan következtetést von le, amilyet a tartalom előállítói szeretnék, hogy levonjon). Ám ha csak a készen kapott tájékoztatásra bízza informáltságát, minden bizonnyal olyan részletekről marad le, amelyek megváltoztathatnák a véleményét. Ami tehát megfigyelhető, hogy elsősorban manipulációs szándék társul e felfogás szerint a keretezéshez. A keretezés pontosabb meghatározása az egyén attitűdjének hagyományos elvárásérték-modelljéből indul ki (Chong–Druckman 2007). Ebben a felfogásban a tárgyhoz való viszonyulás az adott tárgyra vonatkozó meggyőződések súlyozott összessége.

A kognitív nyelvészet keretezési fogalma alapvetően hasonlít Chomsky és Herman modelljéhez (idézi őket Croteau–Hoynes 2000: 241; Bajomi-Lázár 2006), ám paradox módon ebben az esetben is megfigyelhető maga a keretezés: Chomskyék ugyanis kifejezetten a média, nem pedig a nyelv felől közelítették meg a framinget, ezáltal nem magával a „valódi" elmélettel ismertették meg a befogadót, hanem annak „mediatizált" változatával. Ezzel szemben, a kognitív nyelvészet perspektívájából a keretek a világ értelmezésének, felfogásának a megkönnyítését szolgálják; a körülöttünk lévő valóságot keretek, „sematikus interpretációk" alapján dolgozzuk fel (Benczes–Benczes 2018). Bizonyos szavak használatával más és más „keretek" hívhatók életre, különböző reakciókat váltva ki, eltérő végkimenetel felé terelve egy diskurzust. Ez az, amit felhasználnak az újságírók (és a média munkatársai) a hatáskeltés szempontjából. „A (média) keretek nemcsak stabil kognitív reprezentációk, hanem dinamikus, kontextusfüggő tudásstruktúrák is" (Benczes–Benczes 2018: 2). Jelenthetik a hírek konstruálását és feldolgozását, ugyanakkor az egész diskurzust is (Pan–Kosicki 1993; idézi Benczes–Benczes 2018: 2). Összefoglalva tehát, a keretezés minden ember és nyelv szintjén megtalálható, nem kifejezetten a média sajátossága, a média „csak" kihasználja és a saját céljára alakítja a keretezés folyamatát, jellemzően a hatás érdekében. Nemcsak verbális úton lehet ugyanakkor keretezni a „valóságot" (az idézőjel oka, hogy megosztja a tudóstársadalmat, van-e objektív valóság, vagy mindent a saját „keretünkön" keresztül érzékelünk, ettől pedig semmilyen, a világban történő dolgot nem tudunk elvonatkoztatni; Benczes–Benczes 2018: 1).

Technológiával átitatott mindennapjainkban az egyik legfontosabb képesség a multitasking, azaz hogy figyelmünket megosztjuk több cselekvés/tényező között. Valószínűleg ennek egyre elterjedtebbé válása is hozzájárult ahhoz, hogy az emberi figyelemküszöb 12 másodpercről 8 másodpercre esett vissza (Microsoft 2016). A társadalmi átalakulások elvezettek odáig, hogy már nem a föld, az információ vagy a tudás a legértékesebb gazdasági tényező, hanem a figyelem: „akinek nincs, vágyik rá, akinek van, növelni szeretné, befektethető és beváltható, kereshető és kereskedhető" (Aczél–Andok–Bokor 2015: 159). Ennek alapján kijelenthető, hogy az értelmezési keretek kulcsfontosságú szerepet töltenek be attól függetlenül, hogy melyik szakirodalmi nézőpontot alkalmazzák, hiszen a médiafogyasztó gyors információfeldolgozását segítik. Gitlin (1980, idézi: Borah 2011) meghatározását használom a kutatásomban iránymutatásul: véleménye szerint a keretek azok az eszközök, amelyeket az újságírók használnak a hatalmas információmennyiség tömörítésére, közönség elé tárására. Ha pedig ebből a gondolatból indulunk ki, érthető, hogy a deepfake és annak médiareprezentációja nem tekinthető lényegtelennek. Mielőtt azonban ezzel a dilemmával foglalkoznánk, érdemes áttekinteni, mit tartunk deepfake-nek.

2.2. A deepfake alapjai

A deepfake jelenség alapjául a mélytanulás (azaz *deep learning;* innen származik a kifejezés 'deep' része) ellenőrzött formája szolgál: különböző képeket, videókat táplálnak be *(input)* az algoritmusnak felcímkézve *(output),* amelyekből az megtanulja az adott arc legfontosabb karakterjegyeit, jellegzetességeit. A *fake* utótag a 'hamis, ál' szavak angol megfelelőjéből eredeztethető. A szóösszetétel egy Reddit-felhasználónak köszönhetően jött létre 2017-ben (Ajder et al. 2019), ekkorra tehető hivatalosan a deepfake megjelenése. „A deepfake digitálismédia-manipulációt jelent: ultrarealisztikus, a gépi tanulással létrehozott hamis videót, amelynek a szereplői olyan dolgokat tehetnek vagy mondhatnak, amit a valóságban soha nem tennének vagy mondanának" (Dobber et al. 2020: 1 alapján Veszelszki 2021). Azonban a kutatók egy része úgy véli, a szóösszetétel 'deep' tagja a mélytanulás helyett inkább kifinomultsági szintet jelez, és minden olyan technológia idetartozhat, amely ilyen mértékű hitelességet ér el a hamis tartalmak gyártásában (Altuncu et al. 2022). Ennek mintájára megjelent két új kifejezés, a *shallowfake* és a *cheapfake*, amelyek a gyengébb minőségű manipulációra utalnak. Ezen videók előállítóinak bevett módszerei közé tartozik a könnyen hozzáférhető szoftverek használata, amelyekkel lelassítják, felgyorsítják, új kontextusba helyezik a tartalmakat, így érve el a félrevezető hatást (Brady 2020).

Bár gyakran tekintenek a deepfake-re egyszerű arccsereként (angolban *face swapping;* az egyik legelterjedtebb forma), azonban olyan egyéb típusai is előfor-

dulnak, mint az adott arc teljes újraanimálása *(face reanimation)*, bizonyos tárgyak vagy személyek eltüntetése *(object removal)* vagy a szintetikusan megalkotott kép és hang (Reuters 2019; Vincent 2019). Működési elve a mélytanuláson kívül annyival egészítendő ki, hogy a rendszer a minták alapján elkészített saját változatát addig tökéletesíti, amíg az meg nem téveszti a saját ellenőrző szoftverének az érzékelését is, azáltal, hogy minél több vizuális adattal látjuk el az algoritmust (Whittaker et al. 2020). A jelenség veszélyét egyfelől az adja, hogy „a photoshoppolt képekkel ellentétben nemcsak a szemre, hanem a fülre is hat" (Veszelszki 2021); másfelől, hogy manapság már bárkiről gyűjthető elegendő kép és videó a közösségimédia-platformokról, így könnyűszerrel készíthető megtévesztő élethűséggel olyan mozgóképes tartalom is, amelynek megalkotásához az illető nem járul hozzá. Ebbe a kéretlen felhasználási módba ugyanúgy beleértendő a lejáratás vagy a nyereségvágy céljából létrehozott álhírterjesztés, mint a zsarolás és a bosszúpornó (Kirchengast 2020). A hasonló tartalmak előállítása olcsó, nem igényel speciális készségeket, a letiltás pedig az internet korában nem jelent tartós megoldást (Ajder et al. 2019). Bár a médiában többségében a potenciális fenyegetések felől közelítik meg a témát, számos olyan terület akad, ahol a technológia alkalmazása pozitív változásokat hozhat az életünkbe: a bűnüldözés területén a fantomképek elkészítésében és a biztonsági kamerák képeinek javításában jelenthet előrelépést, az oktatásban pedig izgalmasabbá válhat segítségével a tananyagok szemléltetése – akár maga Napóleon is beszámolhat híres hadjáratairól és csatáiról; Mona Lisa mesélheti el a festmény keletkezésének körülményeit; Babits Mihály személyesen mutathatja be a költészetét digitális úton életre keltett felvételek segítségével (Horváth–Mezriczky 2021). Az utóbbi időben azonban a lehetőségek háttérbe szorultak, ennek pedig több oka is van.

2.3. Alvó effektus és Liar's dividend

A deepfake politikai felhasználása nem új keletű jelenség, gondoljunk csak Manoj Tivari esetére Indiában, ami a deepfake első választási kampánybeli alkalmazása volt (Niles 2020): az ellenzéki politikus olyan videókat tett közzé, amelyek kockáról kockára megegyeztek, mégis az egyik felvételen angolul, a másikon pedig a hindi nyelv haryanvi dialektusában beszélt. A mélytanuláson alapuló technológiát használta fel arra, hogy a nyelvileg erőteljesen megosztott Indiában üzeneteivel több állampolgárt érhessen el (végül közel 15 milliót sikerült is). 2022-ben azonban a megváltozott világpolitikai helyzet, az orosz–ukrán háború új lehetőséget teremtett a felhasználók megtévesztésére, hamis információk terjesztésére. Mindkét, hadban álló fél elnökét ábrázolta már félrevezető videó, amelyekben fegyverletételre, illetve békekötésre szólítottak fel (Wakefield 2022; lásd erről részletesen Krasznay Csaba tanulmányát a kötetünkben – *a szerk.*). Ezek jelentőségét két elmélet érzékelteti leginkább.

Hovland és Weiss (1951, idézi Kietzmann et al. 2020) alvó effektusa azt mondja ki, hogy hiába vagyunk tudatában annak, hogy amit látunk tartalomfogyasztás közben, az hamisítvány, hosszú távon annak az üzenetei mégis befolyásolnak minket. Így hiába felismerhető a módosítás az orosz vezető videóján, mégis képes hatást gyakorolni az állampolgárokra. Előfordult olyan eset is, amikor egy hiteles (ez azóta is vita tárgyát képezi) videót hívtak deepfake-nek: a Vlagyimir Putyint ábrázoló klipet, amelyen légikísérők körében beszél a háborúról, az ukrán elnöki hivatal vezetőjének tanácsadója nevezte hamisítványnak, az oroszok pedig tagadták, hogy így lenne (Qubit 2022). Az eset tanulsága tulajdonképpen maga a Liar's dividend (ami egy nehezen magyarosítható kifejezés) elmélet lényegi eleme: a realisztikus manipuláción átesett mozgóképes tartalmak elfogadott szűrésének a hiánya azt vonja magával, hogy már a valóság is hívható hamisítványnak szinte következmények nélkül (Chesney–Citron 2019). Ha ezt összekapcsoljuk a framing-elmélettel és a média szerepével a fogyasztók szemléletének formálásában, kitűnik, hogy a jelenségről alkotott képünk sokszor támpont nélküli. Az igazságszkepticizmus pedig mindenféle tartalom iránt gyanakvást ébreszt bennünk, akár valóságos, akár hamisított tartalmakról legyen szó (Kietzmann et al. 2020). Ezt oldják fel a média által kínált narratívák, keretek, amelyekben nem a fogyasztónak kell döntenie egy-egy videó valóságértékéről. Érdemes tehát megvizsgálni, hogy ezek az értelmezési keretek milyen kontextusba ágyazzák a jelenséget, és hogyan formálják a társadalmi diskurzust.

3. MÓDSZERTAN

A deepfake-et tekintették már a nukleáris fegyverek modern megfelelőjének (Rubio 2018), de a fake news új nehézfegyverének is (Veszelszki 2021); társadalmi megítélése minden túlzás nélkül tekinthető ellentmondásosnak. Kutatásomban ebből kifolyólag arra kerestem a választ, hogy a cikkek milyen témákkal kapcsolják össze a technológiát, és hogyan változik annak online sajtóreprezentációja az első magyar médiamegjelenéstől, 2018-tól a vizsgálat idejére, 2022-re. Azt feltételeztem, hogy a deepfake-videók számának folyamatos emelkedésével összhangban a jelenség egyenletesen növekvő számú publikációt is generál. Emellett a nemzetközi szakirodalom alapján úgy véltem, hogy a legtöbbször deepfake-hez társított témák a pornó, a szórakoztatás és a politika lesznek.

A vizsgálati időszak 2018-tól 2022-ig minden év első félévét fedi. Ennek oka, hogy 2022 jelentős mérföldkőnek számít a deepfake történetében, először használták ugyanis háborús konfliktus során (hamis információ terjesztésére). Ebből kifolyólag fontosnak éreztem, hogy a kutatási minta részét képezze 2022 is. Ugyanakkor, mivel a kutatás végrehajtásának éve a tanulmány írásának pillanatában még nem fejeződött be, az első féléves minták elemzése mellett döntöttem.

Hovland és Weiss (1951, idézi Kietzmann et al. 2020) alvó effektusa azt mondja ki, hogy hiába vagyunk tudatában annak, hogy amit látunk tartalomfogyasztás közben, az hamisítvány, hosszú távon annak az üzenetei mégis befolyásolnak minket. Így hiába felismerhető a módosítás az orosz vezető videóján, mégis képes hatást gyakorolni az állampolgárokra. Előfordult olyan eset is, amikor egy hiteles (ez azóta is vita tárgyát képezi) videót hívtak deepfake-nek: a Vlagyimir Putyint ábrázoló klipet, amelyen légikísérők körében beszél a háborúról, az ukrán elnöki hivatal vezetőjének tanácsadója nevezte hamisítványnak, az oroszok pedig tagadták, hogy így lenne (Qubit 2022). Az eset tanulsága tulajdonképpen maga a Liar's dividend (ami egy nehezen magyarosítható kifejezés) elmélet lényegi eleme: a realisztikus manipuláción átesett mozgóképes tartalmak elfogadott szűrésének a hiánya azt vonja magával, hogy már a valóság is hívható hamisítványnak szinte következmények nélkül (Chesney–Citron 2019). Ha ezt összekapcsoljuk a framing-elmélettel és a média szerepével a fogyasztók szemléletének formálásában, kitűnik, hogy a jelenségről alkotott képünk sokszor támpont nélküli. Az igazságszkepticizmus pedig mindenféle tartalom iránt gyanakvást ébreszt bennünk, akár valóságos, akár hamisított tartalmakról legyen szó (Kietzmann et al. 2020). Ezt oldják fel a média által kínált narratívák, keretek, amelyekben nem a fogyasztónak kell döntenie egy-egy videó valóságértékéről. Érdemes tehát megvizsgálni, hogy ezek az értelmezési keretek milyen kontextusba ágyazzák a jelenséget, és hogyan formálják a társadalmi diskurzust.

3. MÓDSZERTAN

A deepfake-et tekintették már a nukleáris fegyverek modern megfelelőjének (Rubio 2018), de a fake news új nehézfegyverének is (Veszelszki 2021); társadalmi megítélése minden túlzás nélkül tekinthető ellentmondásosnak. Kutatásomban ebből kifolyólag arra kerestem a választ, hogy a cikkek milyen témákkal kapcsolják össze a technológiát, és hogyan változik annak online sajtóreprezentációja az első magyar médiamegjelenéstől, 2018-tól a vizsgálat idejére, 2022-re. Azt feltételeztem, hogy a deepfake-videók számának folyamatos emelkedésével összhangban a jelenség egyenletesen növekvő számú publikációt is generál. Emellett a nemzetközi szakirodalom alapján úgy véltem, hogy a legtöbbször deepfake-hez társított témák a pornó, a szórakoztatás és a politika lesznek.

A vizsgálati időszak 2018-tól 2022-ig minden év első félévét fedi. Ennek oka, hogy 2022 jelentős mérföldkőnek számít a deepfake történetében, először használták ugyanis háborús konfliktus során (hamis információ terjesztésére). Ebből kifolyólag fontosnak éreztem, hogy a kutatási minta részét képezze 2022 is. Ugyanakkor, mivel a kutatás végrehajtásának éve a tanulmány írásának pillanatában még nem fejeződött be, az első féléves minták elemzése mellett döntöttem.

Az említett időszak online sajtómegjelenéseit tartalomelemzéssel vizsgáltam, amelynek során témakörökbe soroltam az elemzési egységeket. Kutatásom során Magyarország piacvezető sajtóelemző szolgáltatását, az IMEDIA-t használtam, amely lehetőséget adott a kulcsszavas keresésre, 2013-ig visszamenően. A korpusz alapját azok a cikkek képezték, amelyek tartalmazták a *„deepfake"* kulcsszót. Ezek közül is kiszűrtem azokat, amelyek blog.hu-s végződéssel rendelkeztek, a blogbejegyzések tényellenőrzése és forrásfelhasználása ugyanis nem feltétlenül rendelkezik azokkal a minőségi kritériumokkal, amelyeknek a vezető médiaorgánumok megfelelnek (ideális esetben). A mintát így összesen 882 darab közlemény adta, ami elegendő mennyiségnek bizonyult a kutatási kérdések megválaszolásához.

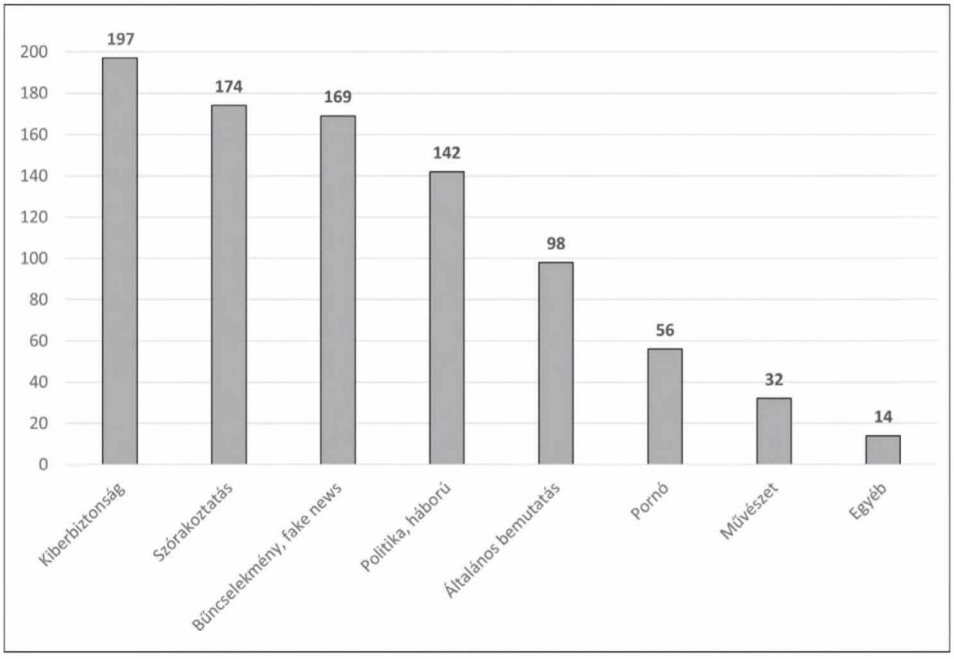
4. EREDMÉNYEK

Általánosságban kijelenthető, hogy az eredmények részben igazolták a kutatási kérdésekhez kapcsolódó hipotéziseket, ugyanakkor meglepetést is okoztak, főként a témákat illetően. Az évek tekintetében jól megfigyelhető egy általános növekedés, amely összefüggésbe hozható a deepfake-videók számának emelkedésével. Egyben az is kitűnik, hogy a kezdeti „fellendülés" után, 2019 és 2021 között szinte ugyanakkora mértékben jelentek meg cikkek a deepfake-kel kapcsolatban, és feltehetően ez a tendencia folytatódott volna akkor is, ha nincs egy, az orosz–ukrán háborúhoz hasonló világpolitikai esemény, amelyben negatív felhasználási módokkal alkalmazzák a technológiát. Ugyanakkor, a jelenség felbukkanása óta kísérte ez a félelem (hogy militáris célokra használják fel) az audiovizuális manipuláció ezen fajtáját számos egyéb aggállyal együtt, így tulajdonképpen csak idő kérdése volt, hogy ez megtörténjen.

4.1. A témák kumulált megjelenéseinek elemzése

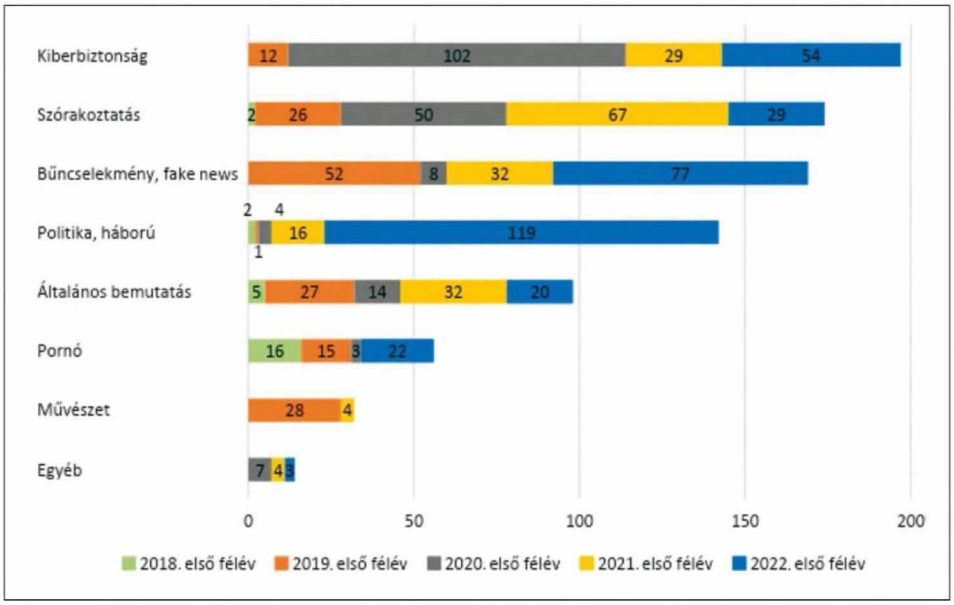
A 882 db online sajtómegjelenés elemzése során 8 db címkét különítettem el, amelyek alkalmasak voltak az elemzési egységek besorolására. Az alábbi csoportok születtek (megjelenésszám szerinti csökkenő sorrendben): *„Kiberbiztonság"*; *„Szórakoztatás"*; *„Politika, háború"*; *„Általános bemutatás"*; *„Pornó"*; *„Művészet"*; *„Egyéb"* (1. ábra). A *„Kiberbiztonság"* kategória foglalta magában azokat a cikkeket, amelyek a deepfake elleni védekezési metódusokat mutatták be; a *„Szórakoztatás"* címke azokhoz az elemekhez kapcsolódott, amelyek egy-egy jól sikerült paródiavideót ismertettek. A konkrét, deepfake-kel elkövetett, pénz- és/vagy adatlopási eseteket a *„Bűncselekmény, fake news"* kategóriába soroltam. Abban az esetben, amikor a cikkek a deepfake politikai felhasználását (kampány, világpolitikai esemény) taglalták, esetleg beszámoltak alkalmazásáról az orosz–ukrán há-

borús helyzetben, a *„Politika, háború"* jelzéssel, ha pedig több terület egyenletes említésével mutatták be a jelenség múltját és jelenét, *„Általános bemutatás"* jelzéssel láttam el őket. A *„Pornó"* kulcsszó szinte magától értetődik, hiszen számos hírességről készítettek a hitelesség látszatát keltő felnőttfilm-jeleneteket, és ezekről az online sajtó is előszeretettel számolt be. Amikor a deepfake-et egy művészi produktum alkotási folyamatába vonták be, az a *„Művészet"* címkét kapta. Az *„Egyéb"* csoport olyan írásokat foglal magában, amelyeket nem tartottam érdemesnek külön kategóriaként kezelni alacsony elemszámuk miatt, hiszen ez elaprózódást eredményezett volna. Ezek tartalmi szempontból a *marketing*hez, *tudomány*hoz és *technológiai innovációk*hoz kapcsolódtak.



1. ábra. A deepfake-témájú megjelenések eloszlása témakörönként (forrás: saját szerkesztés)

Az eredmények alapján kijelenthető, hogy az elemzett időszakban három fő téma határozza meg a deepfake körül kialakult magyar társadalmi diskurzust: a *Kiberbiztonság*, a *Szórakoztatás* és a konkrét csalásokat bemutató *Bűncselekmény, fake news*. Az átlagos felhasználó leggyakrabban azzal találkozik a médiában, hogy milyen védekezési módszerek vannak a deepfake ellen, esetleg melyik nagyvállalat mit tesz a hasonló, manipulált tartalmak visszaszorításáért. Számos videóval lép interakcióba, amelyeken csak az látszik, miként illesztették be például Tom Hollandet Michael J. Fox helyére Marty McFlyként; illetve megismeri azokat az esete-



2. ábra. A témák éves eloszlásának összehasonlítása (forrás: saját szerkesztés)

ket, amelyek során visszaéltek a deepfake-hez kapcsolódó arc- és hangszintetizáló lehetőségekkel.

Jelentős megjelenésszámot tudhat magáénak a *Politika, háború* kategória is, ugyanakkor összesen a 142 db megjelenésből 119 db a 2022-es évben született, az orosz–ukrán háború kapcsán, így nem jelenthető ki, hogy a kiemelt három téma mellett domináns tényezőként lenne jelen a magyar online médiában, a teljes vizsgált időszakban. Magas értéke inkább pillanatnyi képet ad arról, 2022-ben milyen tartalmi kontextus társul leginkább a deepfake-hez (2. ábra). A háborús narratíva nélküli politikai cikkek a választási kampányok manipulálásának eszközeként taglalják a technológiát. Mivel azonban az említett világpolitikai esemény tekinthető az első (nyilvános) alkalomnak, amikor az MI által generált videós tartalmak háborús hadviselési eszközökként szerepeltek, előfordulhat, hogy ez nyitja meg az utat a hasonló alkalmazási módok előtt.

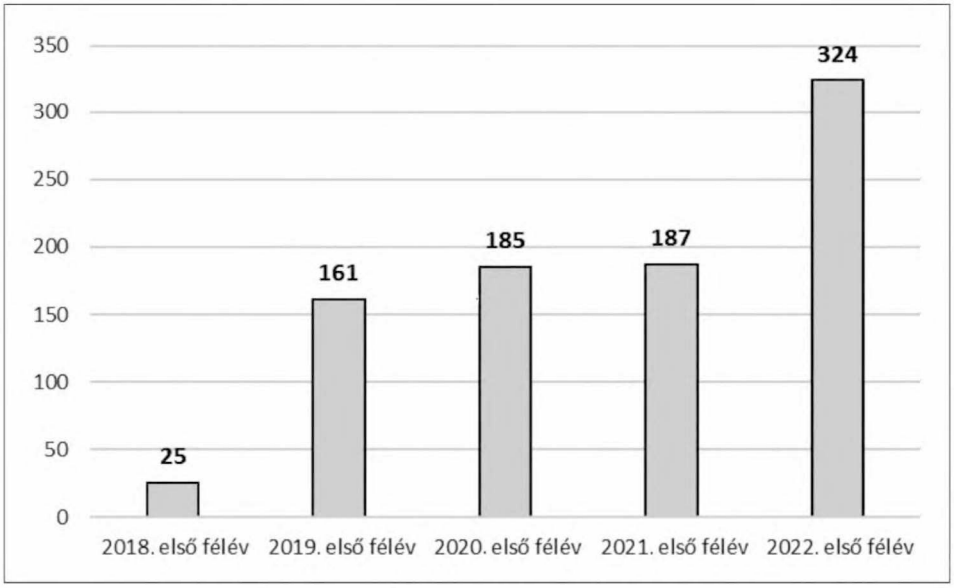
A vizsgált időszakban rendszeresen szerepeltek olyan cikkek is, amelyek azt voltak hivatottak elérni, hogy az olvasó jobban megismerje a deepfake-jelenséget, annak lehetőségeivel és veszélyeivel együtt. Tehát a cikk írója nem foglalt állást semmilyen kérdésben, inkább részletesen és tényszerűen járta körbe azt. Fontos hangsúlyozni, hogy ezek a publikációk azért kerültek külön kategóriába, mert értékítélet nélkül mutatták be a területeket, amelyeken fontos szerep juthat az audiovizuális médiamanipulációnak. Az *„Általános bemutatás”* kategória cikkei-

nek száma alacsony, az összes megjelenésnek mintegy 11%-át teszik ki, azonban vélhetően kielégítik azt az igényt, hogy időről időre bemutassák a jelenség legújabb fejleményeit, kontextusba helyezve azok előzményeivel. A témák tekintetében született sokszínű eredményeket azonban tovább árnyalhatja a megjelenések kvantitatív elemzése.

4.2. Két nagy „robbanás" a megjelenésszámokban

A deepfake-hez kapcsolódó cikkek számában két jelentős ugrás figyelhető meg, 2018-ról 2019-re, valamint 2021-ről 2022-re (3. ábra). 2018 első felében még „csak" 25 darab, míg 2019 hasonló időszakában már 161 darab publikáció említette a *deepfake* kulcsszót, ami magyarázható azzal, hogy a jelenség 2017-ben „debütált", 2018-ra ért el Magyarországra, egy évvel későbbre pedig már kiemelkedő figyelmet generált az online médiában is. Felcímkézett megjelenéseik között is ez látszik: a kezdeti *„Szórakoztatás"*, *„Politika"*, *„Általános bemutatás"*, *„Pornó"* kiegészült a *„Művészet"* és a *„Bűncselekmény, fake news"* kategóriákkal. Megoszlásukat tekintve a *„Pornó"* végzett az élen 2018-ban: a cikkek 64%-a foglalkozott azzal, hogy híresség ekről készült manipulált szexvideók lepték el az internetet. 2019-re megjelentek azok a paródiavideók, amelyekben színészek megjelenését cserélték ki annak megfelelően, melyik másik szerepben látták volna őket szívesen más popkulturális alkotásokban (ez képezi a *„Szórakoztatás"* kategóriát), azonban felbukkantak az első csalások is. Többek között Mark Zuckerbergről, a Meta alapítójáról terjedt el egy felvétel, amelyen a felhasználók személyes adatainak az ellopásáról beszél, mintegy beismerő vallomás gyanánt. 2019 első félévének főbb témái között a meglepetés azonban a *„Művészet":* ekkor adtak ugyanis hírt arról, hogy egy floridai múzeum „feltámasztja" Dalít egy exkluzív tárlat erejéig, és múzeumpedagógiai eszközként fogja használni az MI-alapú videómanipulációt. Ez azt mutatja, hogy a technológia kezdeti megítélése az erőteljesen szexuális tartalmú videók készítésének ellenére sem egyértelműen a veszélyekhez való társítást vonta magával.

A 2021 és 2022 közötti ugrás az eredmények alapján elsősorban főként a háború kitörésével indokolható: hamis híreket terjesztő és politikai csapdák felállítását elősegítő eszközként egyaránt használták a technológiát. Általa megadásra akartak kényszeríteni katonai csapatokat (mint ahogy tette Zelenszkij deepfake-mása egy elhíresült videóban, amelyben az ukrán elnök nevében fegyverletételre szólított fel); a segítséget nyújtó országok politikusainak jó szándékát (többek között Karácsony Gergely budapesti főpolgármestert) használták ki lejáratói célzattal; emellett pedig megjelentek olyan ellentmondásos tartalmak is, mint amilyenekben Vlagyimir Putyin orosz elnök békéről szóló Radnóti-verset szaval. Míg 2021-ben a *„Szórakoztatás"* kategória érte el a legtöbb találatot (a 187 db megjelenésből 67 db-ot,



3. ábra. A megjelenésszámok évek szerinti eloszlása (forrás: saját szerkesztés)

az utána következő *„Bűncselekmény"* és *„Általános bemutatás"* külön-külön ennek már csak nagyjából a felét, 187-ből 32-t), addig 2022-re toronymagasan a *„Bűncselekmény"* (77/324 db) és a *„Háború, politika"* (109/324 db) kategória vezetnek. 2022-re az említett tényezők miatt is (a megnövekedett bűncselekmények és hamis hírek terjesztéséből fakadóan), de felértékelődött a *„Kiberbiztonság"* szerepe: 2021-ben 29 db, 2022-ben már 54 db cikk foglalkozott azzal, hogy bemutassa, miként lehet kiszűrni a manipulált videós tartalmakat, vagy ismertesse egy újonnan megalkotott detektálási metódus részleteit.

Érdemes megjegyezni azonban azt is, hogy a *„Kiberbiztonság"* tekintetében 2020-ban is volt egy erőteljesebb emelkedés, akkor 2019-ről 2020-ra 850%-kal emelkedett azoknak a cikkeknek a száma, amelyek a deepfake elleni védekezést, azok hiányosságait, veszélyeit tárták az olvasó elé. Ennek oka, hogy 2020-ra a jelenség olyan mértékben elterjedtté vált, hogy több nagyvállalat (köztük a Facebook és a Google is) saját detektálóprogram létrehozását tűzte ki célul. A lendület azonban nem tartott ki, és a címkéhez tartozó megjelenések már nem értek fel a 2020-as magasságba: 2020-ban 102 db-ot, 2021-ben és 2022-ben pedig már csak az említett 29 és 54 db találatot generáltak a technológiai megoldásokról szóló híradások. A magyarázat főként abban keresendő, hogy 2020-ban úgy tűnhetett, a szilícium-völgyi összefogás elvezethet a deepfake elleni fellépés csúcsához, egy kézzelfogható produktumhoz, egy algoritmushoz, amely döntő pontossággal ismeri fel a manipulált kép- és videós tartalmakat. Ez azonban nem történt meg,

így a „lelkesedés" is alábbhagyott: számos próbálkozásról számolnak be ugyan a portálok, mégsem jelenthető ki egyikről sem, hogy végső megoldást jelentene az MI-vel szerkesztett videók ellen. Immáron felesleges feltenni a kérdést, hogy kell-e, szükséges-e védekezni a deepfake ellen. A számtalan negatív célú felhasználás és azok megjelenése a médiában – többek között a pornóval, háborúval és a személyiségi jogok ellopásával összefüggésben – látszólag megadja a végső ítéletet a technológiának, és elkönyveli egy társadalomra káros tényezőként.

4.3. Lehetőségek és veszélyek

Érdemes áttekinteni, hogy a nyolc kategória közül melyik milyen narratíva felől közelít a deepfake-hez: A lehetőségeket vagy a veszélyeket látja-e benne? Az *„Általános bemutatás"* kategória semlegesnek tekinthető ebből a szempontból, ugyanis egyszerre jeleníti meg a technológia ellentmondásosságának két pólusát. A *„Kiberbiztonság"*, a *„Bűncselekmény, fake news"*, a *„Pornó"* és a *„Politika, háború"* egyértelműen a veszélyek felől közelítenek: olyan technológiaként mutatják be a deepfake-et, amely ellen védekezni kell; amely személyiségi jogokat sért; amely háborús eszközként szolgál. Az *„Egyéb"* címkéhez tartozó alkategóriák, a *„Tudomány"*, *„Tech"* és *„Marketing"* mellett a *„Művészet"* és a *„Szórakoztatás"* azok, amelyek lehetőségeket mutatnak be azt illetően, hogyan is lehetne a technológia adottságait pozitív célra felhasználni reklámkampányokban, a múzeumpedagógiában vagy a filmgyártásban. Ha ezt a felosztást rávetítjük a megjelenések kvantitatív eloszlására, akkor látható, hogy a lehetőségekhez 220 db, a veszélyekhez 564 db találat kapcsolódik, míg 98 db cikk semleges e kérdésben. Ez pedig elvezethet egy kulcsfontosságú dilemmához: ha a médiában a veszélynarratívával találkozunk döntő többségben, akkor az hatással lehet a lehetőségalapú felhasználás produktumainak befogadására is. Ahhoz azonban, hogy a médiában található pozitív-negatív viszonyulást pontosabb módon lehessen vizsgálni, egy kiegészítő szentimentelemzésre lesz a későbbiekben szükség.

5. KONKLÚZIÓ

A deepfake egyre nagyobb figyelmet kap, ezzel együtt pedig befolyást gyakorol a mindennapokra. Ellentmondásosságát az adja, hogy ugyanaz a technológia képes humoros és szórakoztató, valamint szándékosan félrevezető és destruktív hatásokat is elérni, a tartalom előállítójának céljaitól függően. Veszélyeit számtalan szempontból taglalták már, ugyanakkor kevesebb szó esik azokról a lehetőségekről, amelyeket magával vonhat az alkalmazása. Ha azonban a médiafogyasztók túlnyomó részben a negatív hatásokról és forgatókönyvekről hallanak, kialakulhat

bennük valamiféle belső ellenállás a jelenséggel szemben, függetlenül attól, hogy jó vagy rossz felhasználási móddal találkoznak-e. Ahhoz, hogy erről átfogó képet lehessen kapni magyar viszonylatban, a deepfake kulcsszót tartalmazó magyar online cikkeket vizsgáltam a 2018-tól 2022-ig terjedő időszak első féléveiben. Kutatásom célja volt választ adni azokra a kérdésekre, hogy milyen témákkal kapcsolják össze a médiumok a technológiát, és hogyan változik annak online sajtóreprezentációja az első magyar online sajtószerepléstől e fejezet keletkezésének pillanatáig.

A tartalomelemzés módszerével 882 db publikációt elemeztem, az ennek során kapott eredmények pedig sikeresen választ adnak a fő kérdésekre, a hipotéziseket azonban csak részben igazolják. A deepfake kulcsszót tartalmazó online sajtómegjelenések száma évről évre emelkedik, 2018 vizsgált időszakában még csak 25 db, 2022-ben pedig már 324 db találat született. A legtöbbször előforduló kategóriák a *„Kiberbiztonság"*, a *„Szórakoztatás"* és a *„Bűncselekmény, fake news"*. Ez azt jelenti, hogy az első hipotézis teljesült (hiszen feltételezte, hogy az interneten található, hasonló jellegű videók számának emelkedésével a róluk hírt adó publicisztikák mennyisége is emelkedik), a második azonban nem, ugyanis az a szórakoztatás mellett a pornót és a politikát várta a legnagyobb számú kategóriának a címkék között. A médiában leginkább a védekezési módszerek, a paródiavideók kontextusába ágyazva, valamint olyan esetekhez társítva jelenik meg a deepfake, amelyekben mint a bűnelkövetés eszköze szerepel. A megalkotott nyolc címkéből négy a veszélyek narratívája felől, három a lehetőségeket bemutatva, egy pedig semlegesen tárja az olvasók elé a deepfake jelenséget. Egy következő, kiegészítő kutatásban érdemes lehet a teljes éveket vizsgálni, hiszen az is sokat árnyalhat a képen a médiareprezentáció tekintetében, illetve egy szentimentelemzés elvégzésével kézzelfoghatóbban nyerhető adat a pozitív és negatív kontextusokat tekintve.

A kutatás a választott, tartalomelemzéses módszer alkalmazásából eredően több limitációval is rendelkezik: előfordulhat, hogy számos cikk feldolgozta az audiovizuális médiamanipuláció ezen formáját, mégsem említették benne a *deepfake* kulcsszót, emiatt pedig nem kerültek bele a vizsgálati mintába. Ahogy az is előfordulhatott, hogy – hiába piacvezető médiaelemzési portál – az IMEDIA nem listázott minden előforduló találatot, mert az adott portál – mérete miatt – nem esett bele az algoritmus elemzési körébe. A kapott eredmények ebből fakadóan a vizsgálati mintára igazak, és nem általánosíthatók a teljes évekre vetítve.

A deepfake korai éveiben főként pornóval kapcsolták össze a jelenséget, amelyről sokaknak a felnőttfilmek és a hollywoodi színésznők szinte tökéletes illúziót nyújtó kombinációja jutott eszébe, azonban a 2022-es évre a politika, a háború, a bűncselekmények és a hamis információk szinte megkülönböztethetetlen új formája az, ami a diskurzusokat uralja. Nem csoda tehát, hogy a kiberbiztonság szerepe is felértékelődik, azonban talán ezen a ponton érkezett el a jelenség odáig, hogy érdemes feltenni a kérdést: Ha a pozitív felhasználási módok ilyen csekély

mértékben szerepelnek a médiában, fog-e valaha változni a róla alkotott kép, vagy örökre a digitális atombombát látjuk majd benne? Akármelyik is következik be, az hatással lesz a technológia felhasználási szokásaira is.

SZAKIRODALOM

Aczél Petra – Andok Mónika – Bokor Tamás 2015: *Műveljük a médiát!* Budapest: Wolters Kluwer.

Ajder, Henry – Patrini, Giorgio – Cavalli, Francesco – Cullen, Laurence 2019: The State of Deepfakes: Landscape, Threats, and Impact. *Deeptrace*, október 8. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf [2022. 09. 12.]

Altuncu, Enes – Franqueira, Virginia N. L. – Li, Shujun 2022: *Deepfake: Definitions, Performance Metrics and Standards, Datasets and Benchmarks, and a Meta-Review.* arXiv:2208.10913.

Bajomi-Lázár Péter 2006: Manipulál-e a média? *Médiakutató*, 7/2: 77–95. https://mediakutato.hu/cikk/2006_02_nyar/04_manipulal-e_a_media/ [2022. 09. 12.]

Baudrillard, Jean 2017: A szimulákrum elsőbbsége. *Médiavadász*, május 15. https://mediavadasz.info/jean-baudrillard-a-szimulakrum-elsobbsege/ [2022. 09. 12.]

Benczes, István – Benczes, Réka 2018: From financial support package via rescue aid to bailout: Framing the management of the Greek sovereign debt crisis. *Society and Economy*, 40/3: 431–445.

Borah, Porismita 2011: Conceptual Issues in Framing Theory: A Systematic Examination of a Decade's Literature. *Journal of Communication*, 61/2: 246–263. https://doi.org/10.1111/j.1460-2466.2011.01539.x

Brady, Madeline 2020: Deepfakes: A New Desinformation Threat? *Democracy Reporting International*, július 31. https://democracyreporting.s3.eu-central-1.amazonaws.com/images/20842020-09-01-DRI-deepfake-publication-no-1.pdf [2022. 09. 12.]

Chong, Dennis – Druckman, James N. 2007: Framing Theory. *Annual Review of Political Science*, 10/1: 103–126. https://doi.org/10.1146/annurev.polisci.10.072805.103054

Croteau, David – Hoynes, William 2000: *Media/Society. Industries, Images, and Audiences.* London & New Delhi & Thousand Oaks: Pine Forge Press.

Gitlin, Todd 1980: *The whole world is watching: Mass media in the making & unmaking of the new left*. California: University of California Press.

Kietzmann, Jan – Lee, Linda W. – McCarthy, Ian P. – Kietzmann, Tim C. 2020: Deepfakes: Trick or treat? *Business Horizons*, 63/2: 135–146. https://doi.org/10.1016/j.bushor.2019.11.006

Maras, Marie-Helen – Alexandrou, Alex 2018: Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23/3: 255–262. https://doi.org/10.1177/1365712718807226

Nightingale, Sophie J. – Farid, Hany 2022: AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences* 119/8. https://doi.org/10.1073/pnas.2120481119

Scheufele, Dietram A. 1999: Framing as a Theory of Media Effects. *Journal of Communication*, 49/1: 103–122. https://doi.org/10.1111/j.1460-2466.1999.tb02784.x

Veszelszki Ágnes 2021: deepFAKEnews: Az információmanipuláció új módszerei. In: Balázs László (szerk.): *Digitális kommunikáció és tudatosság.* Budapest: Hungarovox Kiadó. 93–105.

FORRÁSOK

Előd Fruzsina 2022: Már a metaverzumban is elindult a városi területek felértékelődése, több millió dollárért kelnek el központi telkek. *Telex,* február 7. https://telex.hu/zacc/2022/02/07/metaverzum-dzsentrifikalodik-1-2-millio-dollaros-virtualis-ingatlan-tranzakcio-decenterland [2022. 09. 12.]

Microsoft 2016: *A fiatalok nem a szóból, hanem a képből értenek.* Microsoft Magyarország, január 26. https://news.microsoft.com/hu-hu/2016/01/26/a-fiatalok-nem-a-szobol-hanem-a-kepbol-ertenek/ [2022. 09. 12.]

Nilesh, Christopher 2020: We've Just Seen the First Use of Deepfakes in an Indian Election Campaign. *Vice*, február 18. https://www.vice.com/en_in/article/jgedjb/the-first-use-of-deepfakes-in-indian-election-by-bjp [2022. 09. 12.]

Reuters 2019: Identifying and tackling manipulated media. *The Reuters*, december 30. https://www.reuters.com/manipulatedmedia [2022. 09. 12.]

Rubio, Marco 2018: Video: Rubio Discusses Threat "Deep Fake" Technology Poses to U.S. National Security. *rubio.senate.gov,* július 19. https://www.rubio.senate.gov/public/index.cfm/2018/7/video-rubio-discusses-threat-deep-fake-technology-poses-to-u-s-national-security [2022. 09. 12.]

Vincent, James 2019: ThisPersonDoesNotExist.com uses AI to generate endless fake faces. *The Verge*, február 15. https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan [2022. 09. 12.]

Wakefield, Jane 2022: Deepfake presidents used in Russia-Ukraine war. *BBC News,* május 18. Letöltve: https://www.bbc.com/news/technology-60780142 [2022. 09. 12.]

IT ÉS KIBERBIZTONSÁG

KELETI ARTHUR

Nem minden az, aminek látszódni akar – a deepfake és a hitelesség jelene és jövője

A fejezet a deepfake-jelenség megértéséhez ad alapot azon keresztül, hogy megvizsgálja a digitális térben a bizalom, a technológia és az észlelés kapcsolatának hatását a vélt igazságra és az ebből kialakuló hitelességet a befogadó emberben. Feltárja a hitelességgel kapcsolatos informatikai, technológiai és emberi összefüggéseket, és körbejárja a deepfake műszaki és emberi jellegzetességeit. Szót ejt a digitális tartalom és annak befogadásával összefüggő kiberbiztonsági vonatkozásokról. Képet ad a mesterséges intelligencia, a videó- és hanghamisítás és a biztonság viszonyáról, a videó- és hangmanipuláció, illetve a deepfake kialakulásához vezető technológiai lépésekről és lehetőségekről. Végül pedig sorra veszi a mesterséges intelligencia és a deepfake jelenség rossz célú felhasználásának lehetőségeit, egyes eseteit és a védelmi ajánlások, módszerek, megoldások lehetőségeit.

Kulcsszavak: kiberbiztonság, bizalom, hitelesség, AI, védelem

1. ÚT A HITELESSÉG FELÉ – A VÉLT IGAZSÁG, A BIZALOM ÉS A HITELESSÉG KAPCSOLATA

A deepfake alapvetően negatív fogalom. Noha a szó jelentése idővel változhat vagy bővülhet (például a pozitív használatára létrejött kifejezésekkel), de maga az új jelenség maradandónak tűnik. Ahhoz, hogy megértsük, milyen kapcsolatban áll a deepfake a hitelességgel és ezen keresztül az emberi, társadalmi és technológiai biztonsággal, meg kell vizsgálnunk az igazság, a bizalom, az információ viszonyát, valamint értelmeznünk kell a technológia lehetőségeit, képességeit és korlátait.

Úgy tartják, az igazságnak sok oldala van. A nézőpontok, a helyzetek, a társadalmi normáink, sőt még az idő is befolyásolja, hogy mit tartunk igaznak. Amikor igazat adunk valakinek, gyakran az empátiánk vezérel minket, vagyis azon képességünk, hogy együtt tudunk érezni valakivel, bele tudjuk magunkat képzelni a másik helyzetébe. Nem feltétlenül értünk egyet azzal, akinek igazat adunk, de el tudjuk képzelni, hogy az ő szemszögéből, az ő normarendszere szerint igaza van.

Éppen azért, mert sok oldala van, az igazság a tények afféle interpretációja, összerendezése, amelyekről általában úgy vélekedünk, hogy megállapítható róluk, hogy valóságosak-e vagy sem. A filozófia legnagyobb alakjai széleskörűen kutatják, elemzik ezt a kérdést. Eltérő elméletek léteznek arra, hogy a világ tényszerű elemei milyen mértékben szubjektívek. Ahogy Wilhelm Jerusalem (1916: 61–67) fogalmaz, Nietzsche ismeretelméletében határozottan túloz, amikor azt állítja, hogy „az ész igazságtartalma = 0", hogy minden percepciónk csak merő látszat. Jelenségeket és tüneményeket ismerünk meg, de ez a szubjektumtól függetlenül létező realitás, számunkra hozzáférhető és ellenőrizhető megismerés forrásai. Amennyiben elfogadjuk ezt a magyarázatot, akkor tényszerűen állíthatjuk, hogy tegnap esett-e az eső, vagy hogy egy ajtót kinyitottak-e.

Az informatikai rendszereket azért hoztuk létre, hogy ilyen tényeket tároljunk bennük: adatokat, amelyekkel dolgozhatunk, és amelyekből még több adatot állíthatunk elő. Ezért az az érzés alakulhatott ki bennünk, hogy amik az informatikai rendszerekben szerepelnek, azok tények. Sőt, általános megfigyelés szerint bizonyos digitális formátumok (mint például a táblázatok) használata tovább növeli a tényszerűség érzését, különösen szervezetekben dolgozó emberek között (Keleti 2016: 113).

Az informatikai rendszereket előállító és üzemeltető informatikai tudomány és informatikus szakma alapvetően vonzódik az eldöntendő kérdésekhez és az egyszerű válaszokhoz, hiszen szakértői a napi munkájuk során egy bináris világot teremtenek. A programok és a számítógépek nyelvén, a matematika eszközeivel mindent leírhatónak tartanak. A tudósokra pedig felnézünk, és az általános vélekedés már csak ezért is igazat ad nekik.

Tehát amikor általunk valósnak vélt tényeket rögzítünk az informatikai rendszerekben, azok elkezdik önmagukat is igazolni, mivel már maga a médium és annak fenntartói, készítői is támogatják ezt a vélekedést. A digitális adatok pedig nagyon könnyen szaporíthatóak, másolhatóak és akár vég nélkül ismételhetőek, ezért sokkal egyszerűbben kialakulhat belőlük az ismételt igazság illúziója, vagyis az a jelenség, hogy ha valamit kitartóan és hosszan ismétlünk, igaznak kezdjük vélni a tartalmát (Hasher–Goldstein–Toppino 1977: 16, 107–112).

A tények és az igazság vizsgálatának másik perspektíváját a bizalom kérdése jelenti. A bizalom legalább annyira társadalmi jelenség, mint kognitív, a társadalomtudományok legalább olyan sokat foglalkoznak vele, mint az informatikai kutatás. A bizalom definíciója ember és ember között más lehet, mint amikor gépek is belépnek a képbe, akár küldő, akár fogadó oldalon, sőt akkor is, amikor két gép közötti bizalomról beszélünk. Azonban az emberi használatra szánt gépi tevékenységeket (például adatközlés, információk és hírek továbbítása, közlések, tanácsok) nem lehet önmagukban csak társadalmi, pszichológiai vagy informatikai oldalról vizsgálni. Különösen azért nem, mert a technológia fejlődésével a gépek egyre „meggyőzőbben", emberi módon képesek kommunikálni, ennek következ-

tében az emberek viszonyulása is megváltozik hozzájuk. Ezért praktikusnak tűnik a bizalom pszichológiai definícióit és a kommunikációs, informatikai magyarázatait egyszerre figyelembe vennünk, amikor a bizalom kérdését a gépek és emberek viszonylatában vizsgáljuk.

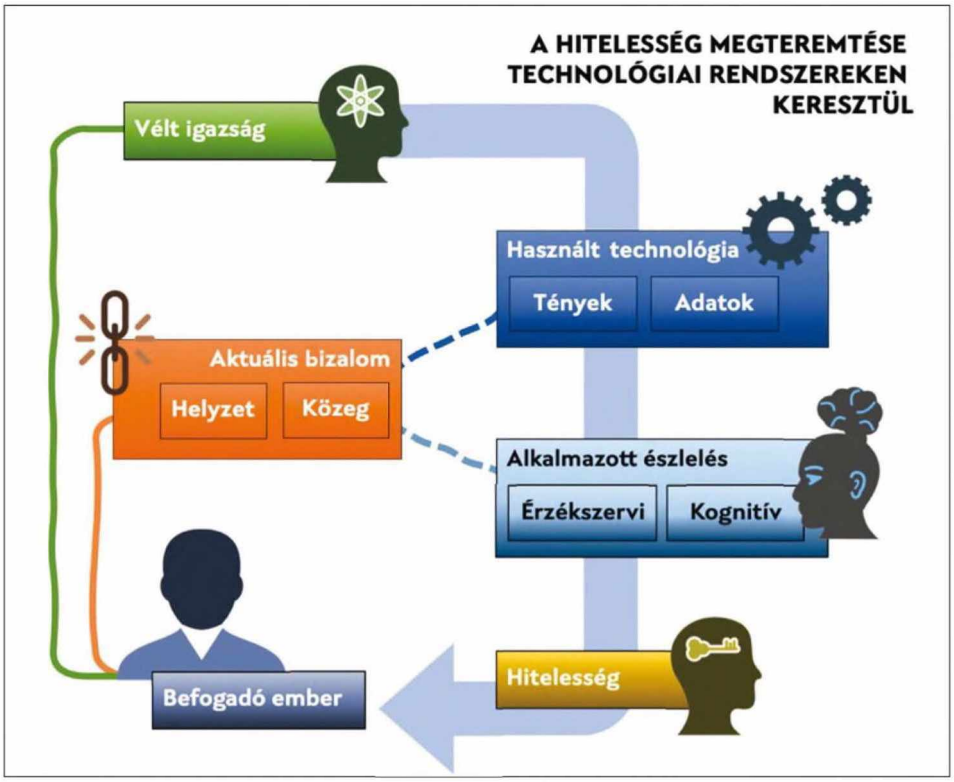
A társadalmi és pszichológiai bizalom az egyén és a társadalom, a kockázat és a bizonytalanság, általános és konkrét esetekre levetített viszonyát vizsgálja. Schilke, Reimann és Cook (2021: 240) a bizalmat úgy határozza meg, mint egy entitás (azaz a megbízó) hajlandóságát arra, hogy egy másik entitással (azaz a megbízottal) szemben kiszolgáltatottá váljon. E kockázat vállalása során a megbízó feltételezi, hogy a megbízott úgy fog cselekedni, hogy a megbízó jólétét szolgálja, annak ellenére, hogy a megbízott cselekedetei kívül esnek a megbízó ellenőrzési körén. Az általános bizalom jellemzően ismeretlenek viszonylag nagy körét, az elemzés társadalmi szintjét és/vagy a tevékenységek széles körét foglalja magában. A szűkített bizalom általában ismerősök viszonylag szűk körére, az elemzés mikroszintjére és/vagy egy konkrét területre vonatkozik. A társadalomkutatás megosztottnak tűnik abban, hogy inkább az általános vagy a szűkített bizalomra fókuszál.

Eközben az informatikai és kommunikációs bizalom kutatói inkább a csatornákra és azok praktikus használatára koncentrálnak. A bizalom ebben az értelmezésben az, ami egy kommunikációs csatornához elengedhetetlen, de a forrástól a célig nem lehet a csatorna segítségével átadni (Gerck 2002: 20). Tehát feltételezhetünk egy külső bizalmi forrást, amely független a csatornán áthaladó tartalomtól.

Az, hogy a bizalmat az egyes tudományágak miként határozzák meg, eltérő lehet, továbbá a bizalom közvetítő technológiai közege, a gépek, valamiféle bináris bizalmat (valós / nem valós) feltételeznek, mégis úgy tűnik, hogy a bizalom maga egyre kevésbé bináris fogalom. A folyamatosan alakuló bizalom jellege változékonyabb lesz, ahogy a technológia beágyazódik az emberi és szervezeti kapcsolatokba (Pew Research Center 2017). Ez pedig ahhoz vezet, hogy a bizalom bináris jellege vagy általános kiterjesztése teljesen megszűnik, eltérő szintjei lesznek, a bizalom erősen függ a kontextustól, és alkalmazkodik a helyzetekhez. A bizalom tehát inkább folyékony, a szituációhoz illeszkedő alakot vesz fel, amelyet a technológia néhol megerősít, néhol pedig éppen elbizonytalanít.

A teljes képhez természetesen hozzátartozik az ember biológiai észlelési képességének kérdése is, hiszen alapvetően abban hiszünk, amit látni vagy tudni vélünk. Éppen ez okozza a deepfake jelenség egyik alapproblémáját, a több millió évnyi evolúció és a gyermekkorban kondicionáltak gyakran mélyebb, ösztönösebb szinten írják felül az elménkben élő, tudott vagy ismert valóságképet. Az érzékszervi észlelésekről egyre gyakrabban tanuljuk meg, hogy becsaphatnak minket.

Így jutunk el a hitelesség kérdéséig, amely a társadalmi és technológiai rendszerek (mint például az internet, még szűkebben pedig a közösségi média, illetve a tartalommegosztó platformok) bizalmi tőkéjét jelentik. A hitelesnek tekintett



1. ábra. A hitelesség megteremtése technológiai rendszereken keresztül (forrás: Keleti 2022)

forrás feloldja a bizonytalanságot, megteremti a bizalmat, befogadóvá teszi az embert (1. ábra).

Tehát *ha a befogadó ember által vélt igazságot a technológia tényekkel és adatokkal támasztja alá egy olyan közegben, helyzetben és észlelésen keresztül, amely felé a befogadó aktuálisan bizalmat érez, akkor hitelesnek tartja majd azt.*

2. INFORMATIKA, TECHNOLÓGIA ÉS EMBER – A DEEPFAKE MŰSZAKI ÉS EMBERI VISZONYAI

Az informatikai és információbiztonság alapja az úgynevezett CIA: a bizalmasság *(confidentiality)*, a sértetlenség *(integrity)* és a rendelkezésre állás *(availability)* hármasa. Ennek forrása az egyik legkorábbi, titoktartással foglalkozó dokumentum, egy 1976-os jelentés volt, amelyet az Egyesült Államok légiereje számára készítettek (United States Air Force 1976: 68, 70). A CIA-modell hármas felosztása első

alkalommal egy NIST- (az amerikai Nemzeti Szabványügyi és Technológiai Intézet) tanulmányban jelent meg, amelyben arról írnak, hogy egy ellenőrnek vizsgálnia kell az adatok bizalmasságát, sértetlenségét és rendelkezésre állását (NIST 1977: 11–13), ez azonban nem bizonyult elégségesnek az informatikai rendszerek védelméhez, ezért a modell tovább bővült (Parker 1998) az úgynevezett parkeri hatossal, amely három elemmel bővítette a CIA-modellt: hitelesség *(authenticity)*, birtoklás vagy irányítás *(availability)*, hasznosság (*utility*, alkalmasság egy konkrét célra).

A hitelesség tehát már a kilencvenes években az információ biztonságának fókuszában volt, különösen azért, mert a felhasználók személyének és az adatok forrásának hitelessége visszatérő problémaként jelentkezett. Nem véletlen, hogy a felhasználók és az adatok hitelesítésének problémájával az 1970-es évektől foglalkozott a brit hírszerzés (GCHQ), és 1976-tól Diffie, Hellman, Rivest, Shamir és Adleman kriptográfusok munkásságának köszönhetően rendszereket építettek a probléma megoldására, amelyeket a mai napig is használnak.

Ugyanakkor az olyan nyilvános informatikai rendszerekben, mint az internet, a hitelesség kérdését még 2020 után sem sikerült teljes mértékben megoldani. Az elméletben már az 1980-as, 1990-es években létező elosztott rendszerek gyakorlati alkalmazására 2008-ban Satoshi Nakamoto név alatt egy ember vagy emberek csoportja hozta létre a blokklánc technológiáját *(blockchain)*, amely az első elosztott, decentralizált blokkláncmegoldás volt. Ennek lényege, hogy a világon bárhol és sok helyen tárolt adatfüzérek, -láncok másolataiban matematikai alapon hitelesített és integritás módban, vagyis módosítás ellen védett információt helyezünk el. Egyszerűen megfogalmazva mindenki birtokolja a láncot, a rajta található eseményekkel együtt, azt meghamisítani, módosítani nem lehet, így ha valaki arra egy bejegyzést tett, akkor annak hitelessége visszavonhatatlanul és mindig ellenőrizhető lesz.

A hitelesség az informatikai rendszerekben a 21. század elejére tehát elért oda, hogy képesek vagyunk az egész bolygóra kiterjedten, időkorlátozás nélkül, széles társadalmi megegyezésnek megfelelően rögzíteni egy esemény megtörténtét. A deepfake-probléma kezelésében ez kulcsfontosságú momentummá válhat, mivel a technológia fejlődésének eredménye kikényszeríti annak *bizonyítását, hogy mikor, mi, hol hangzott el vagy történt meg.* A blokkláncon tárolt hitelesítő információk lehetővé teszik bizonyos hamisítások cáfolatát.

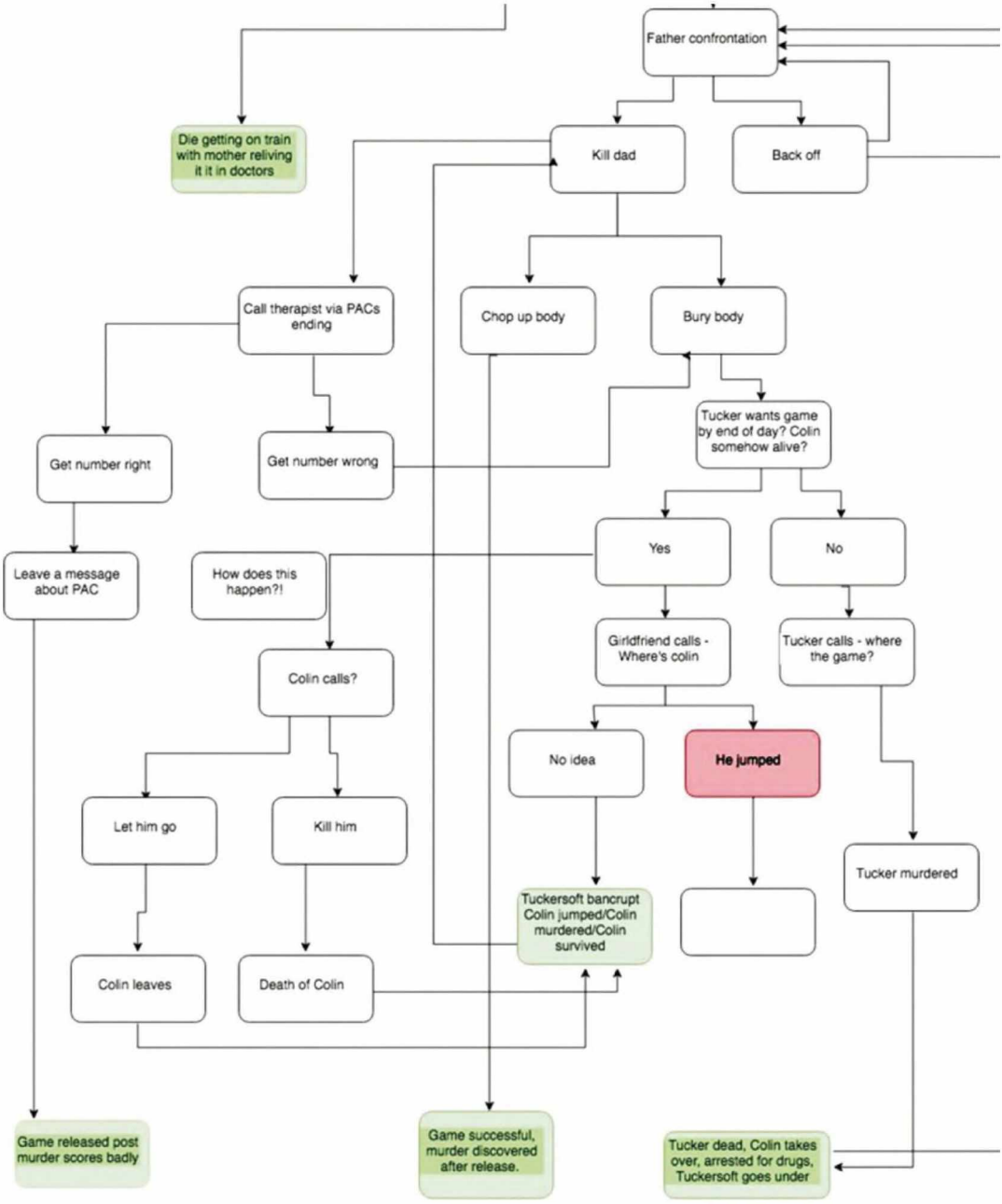
Van azonban egy komoly probléma. Személyes, emberi interakcióinkban a hitelességet nem technológiai megoldásoktól várjuk. A barátaink futóversenyen elért eredményeit maguk a személyek hitelesítik számunkra. A szüleink elmesélt színházélményét nem ellenőrizzük más forrásból. A testvérünk gyerekének első szavairól készült videón nem keressük semmilyen hatóság pecsétjét. A házastársunk hangját a telefonban felismerjük, és nem kérünk hangazonosítást, ahogy a gyerekünktől érkező üzenetben sem vizsgáljuk a hitelességet. Elkerülhetetlen,

hogy ezeket a közléseket, interakciókat egy technológiai rétegen keresztül érzékeljük, ezért a használt technológiába vetett bizalmunk alapvetően határozza meg a közlő és a tartalom hitelességét számunkra.

Ugyanakkor maga a technológia az, amely ezt a bizalmat gyengíteni tudja. A photoshoppolás, vagyis a képek utólagos digitális manipulációja nem volt újdonság senkinek a 2000-es években sem. A fotómodellek mögött „görbülő" csempék kiválóan példázták az emberi test manipulációjának szándékát a divat- és erotikus magazinok fotóin. De ebbe a térbe érkezett egy váratlan játékos: a szűrő. A deepfake feltűnését megelőző, részben azzal párhuzamos jelenség a közösségimédia-platformokon és csevegőalkalmazásokban (mint a Snapchat, az Instagram és később a TikTok) 2011 környékén megjelentek a szűrők, a filterek. Ezek a képmanipulációs eszközök a fotók és videók azonnali megváltoztatásával a mai napig torzítják a valóságot, és egyben felelősek azért is, hogy a felhasználók percepciója saját magukról, a közölt képekről és azok tartalmáról alapjaiban változott meg. A jelenséget tovább erősítette a mesterséges intelligencia támogatta valós idejű arcmanipulációs AR- (kiterjesztett valóság) alapú szűrők alkalmazása, amely egyes kutatások szerint kimutatható hatással van a kozmetikai sebészet elfogadására. Ez azt jelenti, hogy a felhasználók önképének valóságát is alakítja az alkalmazott szűrő. Jonlin és munkatársai (2019) szerint azok a résztvevők, akik konkrét alkalmazások, például a YouTube, a Tinder és a Snapchat fotószűrők használatáról számoltak be, jobban elfogadták a plasztikai sebészetet; más alkalmazások, köztük a WhatsApp és a Photoshop használata pedig szignifikánsan alacsonyabb önértékelési pontszámokkal járt együtt.

A médiafogyasztási szokások átalakulása és a technológia dominanciája ezen a területen is formálja a nézők, fogyasztók gondolkodását. Olyan megoldások, mint a Netflix *Fekete tükör – Interaktív* (Black Mirror: Bandersnatch) film a néző által választható, alakítható eseményei teljes megoldási diagramot is kitesznek, amelyen követhető a film összes lehetséges befejezése (Hurley–Leon 2022; 2. ábra), és ezzel alakítják a nézőnek a fogyasztott médiatartalomról alkotott képét. Emellett pedig a digitális technológiával és a deepfake alapjául szolgáló gépi tanulással „feltámasztott" vagy megfiatalított színészek, énekesek jelenléte a modern médiában is megszokottá kezd válni. *A 21. századi ember egyre kevésbé gondolja egy digitális tartalomról, hogy az valós, állandó és megváltoztathatatlan.*

Mégis igaz, hogy a technológia képességeivel egyre inkább tisztában lévő digitális médiafogyasztó továbbra sem feltétlenül akar állandóan kételkedni, forrásokat keresni és hitelesíteni. Ennek összetett, pszichológiai és technológiai okai is vannak. Ilyen például a heurisztikus, próbálkozó-tapasztalati megközelítés és a döntések leegyszerűsítése. Az emberi elme nem véletlenül hatékony: nem gondolja túl a dolgokat. Ez egyben azonban azt is jelenti, hogy amikor megvizsgálunk valamit, nem vesszük figyelembe a tényezők teljességét, például amikor az interneten olvasott vagy látott dolgok hitelességét mérlegeljük.



2. ábra. A Netflix Bandersnatch folyamati diagram részlete
(forrás: Hurley 2022)

Mint kognitívan fejlett szervezetek, az egyszerűsítés érdekében a lehető legkönnyebben kivitelezhető stratégiákat alkalmazzuk. A forrás hitelesítésében vagy az információkeresésben részt vevő internetfelhasználók valószínűleg egyszerű heurisztikákat fejlesztenek ki, hogy egyszerű számítással megvalósítható módon válasszák ki a megbízható információforrásokat, és annak eldöntésében, mit hisznek el és mit nem (Taraborelli 2008: 194). A „kényelmes igazság", vagyis a könnyen, kevés gondolkodással elfogadható vélt igazság elfogadása alapvető emberi motívumnak tűnik, és az érvelés, gondolkodás iránti alulmotiváltság egyike a kulcsfaktoroknak abban is, miért sérülékenyek az emberek az álhírekkel szemben (Pennycook – Rand 2019).

3. A MESTERSÉGES INTELLIGENCIA, A VIDEÓ ÉS HANGHAMISÍTÁS ÉS A BIZTONSÁG VISZONYA

Tehát a deepfake manipulált videói vagy audiotartalmai olyan közegbe érkeznek, amelyben az emberek digitális hitelességhez fűződő viszonya eleve komplikált. A manipuláció lényege az egyén megtévesztése és ehhez minden rendelkezésre álló technikai eszköz alkalmazása. Nem a tökéletesség a cél, elég, ha a manipulált tartalom már „kényelmesen" befogadható – a hitelesítést megoldja a befogadó ember technológiába és érzékszerveibe vetett bizalma, valamint a kényelmes hozzáállása.

Annak érdekében, hogy a megfelelően megtévesztő tartalmat elő tudja állítani, a deepfake a tág értelemben vett mesterséges intelligencia (AI) kategóriájába tartozó mélytanulás vagy gépi tanulás területét használja. Mára a technológia széles körben elérhető, éppen ez adja a legnagyobb kiberbiztonsági veszélyfaktort. Mint a legtöbb hétköznapi alkalmazás, a képekkel és videókkal összefüggő mélytanulás-alapú manipulációk is gyakran egyetemi kutatásokból alakultak ki. Ha arra vagyunk kíváncsiak, hogy milyen deepfake-technológiával próbálják majd a kiberbűnözők becsapni az áldozataikat néhány éven belül, akkor érdemes áttekintenünk a friss egyetemi és vezető technológiai cégek AI és gépi tanulással összefüggő publikációit.

Nagyjából elmondható, hogy egy működő modellel demonstrált videó- és hangmanipuláló vagy -generáló technológia a kutatói bemutatót követően egy éven belül tökéletesedik. Két vagy három követő verzióban kijavítják a hibáit, és széles körű alkalmazása egy-két éven belül a nem kutatói közönség számára is elérhető lesz. Az is egyre gyakrabban fordul elő, hogy a kutatási versenyben szereplő résztvevők (akár egyetemi, akár technológiai cégek) a működő modellek demonstrációja végett hozzáférhetővé teszik az algoritmusokat. A GPT3 modellt, ezt a különösen jelentős, nagy nyelvi modellt, amely több AI-alapú megoldás alapja lett, 2019-es megjelenésekor készítője, az OpenAI (egy nonprofit és azóta már

piaci kutató laboratórium is) nem tette nyilvánosan használhatóvá, és ezzel nagy felháborodást váltott ki a kutatókból és a felhasználókból. A nyilvánosság korlátozására az OpenAI kutatói szerint részben éppen azért volt szükség, hogy elkerüljék a rosszindulatú felhasználást, és ellenőrzés alatt tudják tartani, hogy a technológiához hozzáférő szakértők mit csinálnak vele.

A 2017-es Synthesizing Obama projekt után alig egy évvel olyan, a széles közönség számára is elérhető alkalmazások jelentek meg, mint a DeepFaceLab. Ezt követően már a hétköznapi használatban is lehetett találkozni olyan projektekkel, mint a CTRL Shift Face (2019), amely Bill Hader parodisztikus improvizációjára vetítette rá Arnold Schwarzenegger arcát. A videó 2022 őszén már húszmilliós megtekintésnél járt a YouTube-on.

Zakharov és munkatársai 2019 nyarán publikálták azt a bemutatót, amelyben egymással versenyző számítógépekkel építették fel és szólaltatták meg a festmény Mona Lisa arcát. A technológiáról készült videó megdöbbentő, látványos és gondolatébresztő. A kutatók egészen új utakat találtak egy forrást adó valódi emberi arcmodell gesztusainak átültetésére egy fotóból vagy festményből képzett arcra. A kutatás vezetőjével, Egor Zakharovval folytatott beszélgetéseimből azonban már akkor komoly tanulságokat lehetett levonni a technológia biztonságra gyakorolt hatásáról. A használt eszközök pontos működéséről, vagyis arról, hogy részletesen és pontosan miként állítja elő a tanuló algoritmus az eredményt, a kutatók maguk sem tudnak sokat. A tanuló algoritmust ismerjük, de hogy pontosan miért jut arra, amire jut, már kevésbé vagy egyáltalán nincsen tudásunk. A műszaki háttér ilyen jellegű bizonytalansága az összes mesterségesintelligencia-érintettségű projektre érvényes; ez egyben egyike azoknak a biztonsági veszélyfaktoroknak, amelyeket kezelnünk kell.

Az Európai Unió abban látja a helyzet megoldását, hogy nagyobb transzparenciát és az AI elmagyarázhatóságát követeli meg új rendeleti keretrendszerében (European Commission 2021: 2–4). Ennek értelmében bizonyos mesterségesintelligencia-rendszerek esetében különleges átláthatósági követelményeket írnak elő, különösen, ha egyértelműen fennáll a manipuláció veszélye (például csetbotok használata során). A felhasználóknak tisztában kell lenniük azzal, hogy egy géppel lépnek kapcsolatba.

Ezeket a szabályokat a deepfake-technológiával előállított (vagy folyamatosan vezérelt) AI-ok esetében is alkalmazni kell, ezzel csökkentve a szolgáltatást használó vagy a mesterséges intelligenciával interakcióban lévő ember kockázatát. Ebből az is következik, hogy a jövőben a Deepfake/Deepvoice vagy egyéb AI-alapú hamisítást, megtévesztést használók majd az EU és a tagországok szabályait sértik meg.

Az AI hanghamisítási képességének lehetősége már a 2010-es évek közepétől elérhető. Többek között a WaveNet ötletének felvetésével indultak meg a fejlesztések a gépi tanulás területén. A WaveNet készítői egy mély neurális hálózatot mutattak be, amely képes feldolgozatlan hanghullámformák létrehozására. A mo-

dell gyorsan tanult a másodpercenként több tízezer hangmintából álló adatokon, amelyekből olyan hangmintákat állított elő, amelyeket az emberi hallgatóság sokkal természetesebb hangzásúnak ítélt, mint a korábbi rendszerek által készített anyagokat. Mindezt ráadásul angol és mandarin nyelven is képes volt elvégezni. A WaveNet több beszélő jellemzőit is hasonló pontossággal tudta megragadni, és a beszélő azonosítása alapján képes volt váltani is közöttük. Meglepő volt és egyben előre jelezte a mélytanuló algoritmusok univerzalitását, hogy amikor zenemodellezésre tanították, eredeti és gyakran rendkívül valósághű zenei darabokat hozott létre (van den Oord et al. 2016).

Ezek a kezdeti, ám nagyon élethű hangszintetizáló jellemzők fejlődtek tovább a Deepvoice-ban, amellyel kapcsolatban egy 2017-es publikációban már a Wavenethez képest négyszázszoros (!) sebességnövekedésről számoltak be (Arik et al. 2017), mindössze öt hónappal a korábbi technológia megjelenése után. Ez kiválóan példázza az AI és a tanuló algoritmusok fejlődésének képességét. Ezt követték olyan megoldások a TTS (a *text-to-speech*, a szövegből hangkészítés) területén, amelyek már a leírt szöveget képesek akár több beszélő hangján élethűen szintetizálni (Giuseppe et al. 2021). Ezzel érkezünk meg a hangklónozás (hangmásolás, *voice cloning*) területéhez és ennek valós idejű megvalósításához (Jia et al. 2018), továbbá az SV2TTS képességekhez *(speaker verification to multispeaker text-to-speech)*, vagyis a beszélő azonosításától a több beszélő szövegéből hangképzés témaköréhez.

A védelmi szakemberek biztosan számolhatnak egy hamisítási technológia gyors tökéletesedésével akár néhány hónap alatt. Ez egyben azt is jelenti, hogy *a kutatók által kifejlesztett legfrissebb technológiák a deepfake területén, legyen szó hangról vagy képről, nagy eséllyel tökéletesednek egy éven belül, és egy-két éven belül hétköznapi használatba is kerülhetnek.*

Arra is számítanunk kell, hogy a kiberbűnözők olcsón és hatékonyan tudják bevetni az ilyen jellegű technológiákat. Amint az meg is történt egy 2021-es bírósági dokumentum nyomán nyilvánosságra került 2020-as csalási esetben (Brewster 2021). Egy hongkongi bank tisztviselőjét tévesztették meg egyik üzletfelük hangjának másolásával *(voice cloning)*, és vették rá a bankot egy 35 millió dolláros utalás teljesítésére (a hírekben korábban egy dubai bank szerepelt, de ez téves információnak bizonyult). Ez akkoriban a második eset volt egy 240 000 dolláros angliai csalás után, amely hasonló technika használatával, az ügyvezető hangjának másolásával történt még 2019-ben (Stupp 2019). Utóbbi esetben a megtévesztett áldozat úgy nyilatkozott, hogy még a lemásolt személy nyelvhasználatára jellemző német akcentust is hallotta a megtévesztő, másolt hangban, végső soron ez is segítette a hitelesség látszatának megteremtését.

4. A DEEPFAKE ÉS AZ AI-ALAPÚ MANIPULÁCIÓ VESZÉLYEI ÉS MEGOLDÁSAI

Az általános használatban elérhető gépi tanulás a képgenerálás és manipuláció terébe nagyjából 2014-ben érkezett meg, amikor Ian Goodfellow és kollégái egy Generative Adversarial Net (GAN) modellt javasoltak, vagyis egymással versengő hálózati megoldást, amely képeket állít elő, és versenyezve egyre jobb eredményre készteti magát. A javasolt GAN keretrendszerben a generatív modell egy ellenféllel áll szemben: egy diszkriminatív modellel, amely megtanulja meghatározni, hogy egy minta a modelleloszlásból vagy az adateloszlásból származik-e. A GAN olyan, mint egy hamisítókból álló csapat, akik hamis pénzt próbálnak előállítani és észrevétlenül felhasználni, míg a vele versengő diszkriminatív modell a rendőrségre hasonlít, amely megpróbálja észrevenni a hamis pénzt. Ebben a játékban a verseny mindkét csapatot arra ösztönzi, hogy addig fejlesszék módszereiket, amíg a hamisítványok megkülönböztethetetlenek lesznek a valódiaktól (Goodfellow 2014: 1–2). Erről a fejlesztők többek között a Neural Information Processing Systems 2016 (NIPS 2016) konferencián számoltak be nagy sikerrel, és az azt követő egyre fejlettebb modellek már a biztonsági szakértők figyelmét is felkeltették, mivel a megoldás előrevetítette, hogy az ilyen rendszerek automatikusan képesek lesznek szinte tökéletes képhamisítások elkészítésére. Az ezt követő években a kép-, a videó- és a hanggenerálás egymással párhuzamosan fejlődött, de hasonló alapokról építkezve. Amikor a technológiák egymáshoz egyre közelebb értek, olyan eredményeket produkáltak, mint a 2018-as Barack Obama-deepfake (Peele, BuzzFeedVideo 2018).

A többi (nem videó- vagy hanggeneráló) AI-technológia megjelenésével a védelmi területen dolgozó szakértők és kutatók (Brundage et al. 2018: 10) idejekorán figyelmeztettek arra, hogy a mesterséges intelligencia rosszindulatú felhasználása veszélyeztetheti a digitális biztonságot (például azáltal, hogy bűnözők gépeket képeznek ki arra, hogy emberi vagy emberfeletti teljesítményt nyújtva feltörjék az áldozatok eszközeit, fiókjait, vagy manipulálják az áldozatokat), a fizikai biztonságot (például nem állami szereplők felfegyverzik a piaci drónokat) és a politikai biztonságot (például a magánszféra védelmével figyelmen kívül hagyó megfigyelés, profilalkotás és nyomásgyakorlás, esetleg automatizált és célzott dezinformációs kampányok révén).

Ezért egyre nagyobb figyelemmel kell követni a leírt szövegből videó- és hangmanipulációs eredményeket felmutató AI-kutatási eredményeket, amelyek pozitív hozadékai a filmipar, a telekonferencia, az oktatás, a videóprodukciók területén óriási fejlődést hozhatnak, de pontosan ugyanezen képességük biztonsági szempontból veszélyessé is teszi őket. Ilyen például a beszélőfejek iteratív szövegalapú szerkesztése neurális retargeting segítségével, amellyel egy már létező videóban egy beírt szöveg megadásával mind a videó-, mind a hangtartalom módosítható.

Ily módon az alany nem azt mondja, amit eredetileg mondott, hanem azt, amit beírtunk helyette. Ez a technológia olyan iteratív beszélőfejes videószerkesztő eszközt mutat be, amely ismert, jó minőségű szintézisek technikájára épül, viszont jelentősen csökkenti a manipulált videó elkészítéséhez szükséges időt, több óráról körülbelül 40 másodpercre egy hatszavas szerkesztés esetén. Emellett egy gyors ajakmozgás-kereső megoldásnak köszönhetően csökkenti a célszereplő, vagyis a mintavideó adatigényét, amelyhez így akár két-három perces minta elég lehet. A rendszer önfelügyelt neurális retargeting-technológiát használ az ajakmozgások célszereplőre való átvitelére. Sőt az eredmények finomításához is biztosítanak lehetőségeket, például elsimíthatják az ugrásszerű átmeneteket, kikényszeríthetik a száj zárását, és lehetővé teszik a nem csupán a beszédet kísérő szájmozdulatok beillesztését (Yao et al. 2020: 2).

Az ilyen megoldások rohamos fejlődésének hatására már korábban is felmerült, hogy a mesterséges intelligencia által megvalósítható megtévesztés ellen több rétegben kell fellépni. Egyrészt a politikai döntéshozóknak szorosan együtt kell működniük a műszaki kutatókkal a mesterséges intelligencia lehetséges rosszindulatú felhasználásának vizsgálata, megelőzése és mérséklése érdekében. Másrészt a mesterséges intelligencia kutatóinak és mérnökeinek komolyan kell venniük a munkájuk kettős felhasználású jellegét, lehetővé téve, hogy a visszaélésekkel kapcsolatos megfontolások befolyásolják a kutatási prioritásokat és normákat, és proaktívan forduljanak az érintett szereplőkhöz, ha a káros alkalmazások előre láthatóak. Továbbá a kettős felhasználással kapcsolatos aggályok kezelésére kiforrottabb módszerekkel rendelkező kutatási területeken – például a kiberbiztonság területén – meg kell határozni a legjobb gyakorlatokat, és azokat a mesterséges intelligencia esetében is alkalmazni kell. Végezetül aktívan törekedni kell arra, hogy az érdekeltek és a szakterületi szakértők körét bővítsük e kihívások megvitatásában (Brundage et al. 2018: 52).

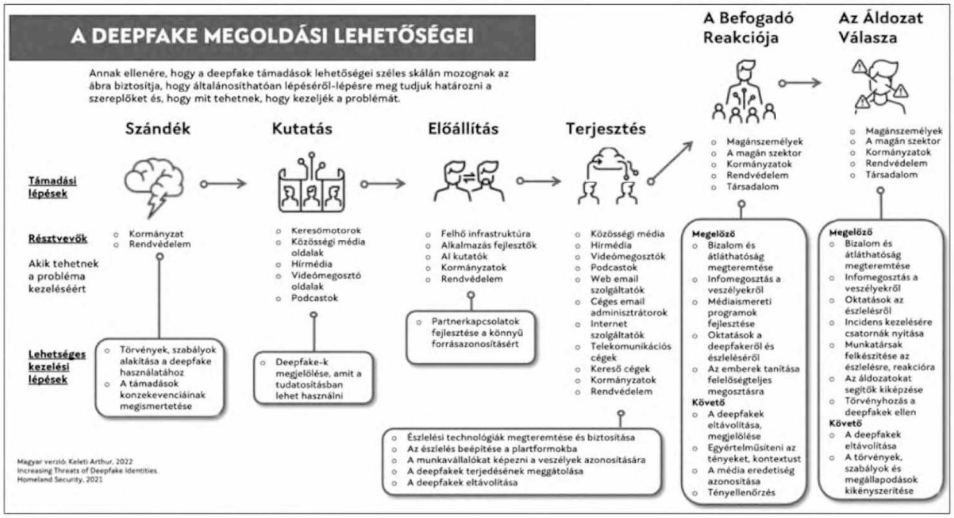
A mesterséges intelligenciával kapcsolatos bűnüldözési, rendvédelmi, kiberbiztonsági és más területek együttes vizsgálatával megjelent az AIC (AI crime), vagyis a MI-bűnözés, a mesterséges intelligencia segítségével elkövetett bűntények, támadások, szabálysértések területe. Az AIC elméleti alapjait és létrejöttét a közösségi média felhasználóit célzó csalások automatizálását bemutató kísérletek és esetek nyilvánosságra kerülése tette lehetővé. Mivel azonban az AIC még mindig viszonylag fiatal és eredendően interdiszciplináris terület – a társadalmi-jogi tanulmányoktól a formális technológia tudományáig terjed –, nem tudhatjuk pontosan, merre halad, mi lesz a jövője. Az AIC terét és lehetőségeit növeli, hogy a kiberbiztonsági szférán belül a mesterséges intelligencia rosszindulatú és támadó szerepet tölt be – párhuzamosan a defenzív AI-rendszerek fejlesztésével és telepítésével, amelyeket a cégek, kormányok, szervezetek azért vezetnek be, hogy növeljék ellenálló képességüket a támadások elviselésében és robusztusságukat a

támadások elhárításában, valamint a felmerülő fenyegetésekkel szemben (King et al. 2019–2020: 1, 114).

A tényellenőrző szervezetek – amelyek a mai napig fontos szerepet töltenek be a hírek ellenőrzésében – a deepfake-kel nehezen boldogulnak, a technológia összetettsége és a jelenség gyorsasága miatt. Ezért a reaktív, manuális megoldások helyett a deepfake elleni védelem kettős rétegben valósítható meg. Az első rétegben a modellrendszerekben le kell fednünk a teljes életciklusokat a kutatástól az innováción át egészen az áldozatok védelméig. A második, technológiai rétegben gondoskodnunk kell biztonsági célú alkalmazások kifejlesztéséről, elterjesztéséről és használatáról. Ilyen például a Resemblyzer, egy szabadon elérhető megoldás, amely segít összehasonlítani a hangforrásokat, és egy 256-os értékelem segítségével igyekszik megkönnyíteni a hamisítás lehetséges azonosítását. A hang frekvenciájának, tónusának valódiságát vagy éppen az élő arc meghatározását olyan technológiák támogatásával kell előremozdítanunk, mint az ID R&D, amely egyben azonosító feladatokat is el tud látni. Hasonló megoldások keresését tűzte ki célul az amerikai védelmi minisztérium DARPA MediFor/SemaFor (Media Forensics / Semantic Forensics) projektje is, amely technológiai szervezetek összefogásának eredményeivel járul hozzá a küzdelemhez (Corvey 2019).

A közösségi média, amely a videómanipulációs anyagok közlésének elsődleges forrása, több fronton küzd a deepfake negatív hatásai ellen. Tájékoztató kampányokat készítenek, jelentéseket adnak, de a rendszerek aktív védelmével és hangolásával is igyekeznek tenni az áldozatok védelméért. Ennek előremutató példája a 2019-ben létrehozott Deepfake Detection Challenge (vagyis a deepfake-észlelési verseny), amely a Facebook (azóta Meta), az Amazon Web Services (AWS), Microsoft partnerség az AI-ért, a Microsoft és a Cornell Tech, az MIT, az Oxfordi Egyetem, a UC Berkeley, a Marylandi Egyetem, College Park és a New York-i Albany Állami Egyetem munkatársainak közreműködésével indult el, és azt célozta meg, hogy a legjobb megoldások kaphassanak teret és figyelmet, amelyekkel a manipulációt szűrjük.

A technológia mellett azonban a védelem elsődleges megoldása mégis a folyamatokban, a résztvevők minél nagyobb arányú bevonásában rejlik. A deepfake összetettsége és kiszámíthatatlansága miatt az enyhítő intézkedéseknek széles körűnek kell lenniük. A rendelkezésre álló emberközpontú és technológiai megoldások lehető legszélesebb körét kell felhasználnunk (Brooks et al. 2021: 28).



3. ábra. A deepfake probléma megoldási lehetőségei (Keleti 2022) (Increasing Threats of Deepfake Identities, DHS, 2021)

A védelem technológiai területén további megfontolásra ajánlhatóak olyan terjedőben lévő megoldások, mint a blokklánc, amely hitelesíthet egy megtörtént eseményt, annak tartalmát, a videó vagy hang lenyomatát, ezzel bizonyítva, hogy az adott cselekmény pontosan mikor, hol, milyen körülmények között történt, és mi volt a tartalma.

5. ÖSSZEGZÉS

A fejezetet azzal kezdtük, hogy a hitelesség és a bizalom témakörét jártuk körbe. Eljutottunk odáig, hogy az AI nemcsak keletkezteti a problémákat, hanem a megoldásban is alapvető szerepe van. Ezért különösen fontos, hogy tisztázzuk a viszonyunkat ezzel az ellentmondásos technológiával, és minél mélyebb bizalmat építsünk ki a deepfake elleni küzdelemben nélkülözhetetlen AI-megoldások iránt. Ehhez ad segítséget az amerikai szabványügyi testület, a NIST ajánlása, amely szerint a mesterséges intelligenciába vetett bizalom attól függ, hogy az emberi felhasználó miként érzékeli a rendszert. Ha az AI-rendszer magas szintű technikai megbízhatósággal rendelkezik, és a megbízhatósági jellemzők értékeit a felhasználási kontextushoz és különösen az adott kontextusban rejlő kockázathoz elég jónak ítélik, akkor az AI-t felhasználó befogadó ember bizalmi szintje emelkedik, a bizalmi viszony kialakulásának a valószínűsége megnő. Ez a felhasználói megíté-

lésen alapuló bizalom az, amelyre az ember és az AI közötti együttműködés során szükség lesz (Stanton–Theodore 2020: 19).

A fejezet nem tárgyalta a deepfake jelenségen kívül eső, de azzal rokon tényezőket, többek között a kiterjesztett és virtuális valóságokat, amelyek alapjaiban befolyásolják azt, ahogyan a körülöttünk létező világot érzékeljük. A 21. század technológiai fejlesztései alapvetően teszik majd próbára az érzékelésünkbe, a vélt tudásunkba vetett bizalmat, ezért a lehető legnagyobb rugalmassággal és nyitottsággal szükséges viszonyulnunk minden olyan élethelyzethez, amelyet a technológia befolyásol majd. Reményem szerint ez a fejezet is hozzájárul majd ahhoz, hogy megfelelően válasszuk meg a bizalom szintjét, amikor tartalmakat fogadunk be, és reagálunk a kibertérben látottakra és hallottakra.

SZAKIRODALOM

Brooks, Tina et al. 2021: Increasing Threats of Deepfake Identities. *Homeland Security*, szeptember 14. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf [2022. 10. 12.]

Brundage, Miles – Avin, Shahar – Clark, Jack – Toner, Helen – Eckersley, Peter – Garfinkel, Ben – Dafoe, Allan – Scharre, Paul – Zeitzoff, Thomas – Filar, Bobby – Anderson, Hyrum – Roff, Heather – C. Allen, Gregory – Steinhart, Jacob – Flynn, Carrick – Ó hÉigeartaigh, Seán – Beard, Simon – Belfield, Haydn – Farquhar, Sebastian – Lyle, Clare – Crootof, Rebecca – Evans, Owain – Page, Michael – Bryson, Joanna – Yampolskiy, Roman – Amodei, Dario 2018: The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *ArXiv*, február. 10., 52. https://arxiv.org/pdf/1802.07228.pdf [2022. 10. 12.]

Chen, Jonlin – Ishii, Masaru – Bater, Kristin L. – Darrach, Halley – Liao, David – Huynh, Pauline P. – Reh, Isabel P. – Nellis, Jason C. – Kumar, Anisha R. – Ishii, Lisa E. 2019: Association Between the Use of Social Media and Photograph Editing Applications, Self-esteem, and Cosmetic Surgery Acceptance. *JAMA Facial Plastic Surgery*, 21/5: 361–367. https://doi.org/10.1001/jamafacial.2019.0328 [2022. 10. 01.]

Corvey, William 2019: *Semantic Forensics (SemaFor).* Dept of Defense, Defense Advanced Research Projects Agency (Darpa), november 19. https://www.darpa.mil/program/semantic-forensics [2022. 10. 11.], https://sam.gov/opp/a8883be78ac1442e8a22924011fc13c4/view [2022. 10. 11.]

European Commission 2021: New rules for Artificial Intelligence – Questions and Answers. Press corner, április 21., 2–4. https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683, https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_21_1683/QANDA_21_1683_EN.pdf [2022. 10. 09.]

Gerck, Ed 2002: Trust as Qualified Reliance on Information. Part I, Technical Report. *The COOK Report on Internet*, X/10: 19–24.

Goodfellow, Ian J. – Pouget-Abadie, Jean – Mirza, Mehdi – Xu, Bing – Warde-Farley, David – Ozair, Sherjil – Courville, Aaron – Bengio, Yoshua 2014: *Generative Adversarial Networks.* Departement d'informatique et de recherche operationnelle Universite de Montreal, június 10., 1–2. https://arxiv.org/abs/1406.2661 [2022. 10. 11.]

Hasher, L. – Goldstein, D. – Toppino, T. 1977: Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16: 107–112.

Hurley, Leon 2022: Every Black Mirror: Bandersnatch ending and how to get them. *Flowchart*, gamesradar.com, április 21. https://cdn.mos.cms.futurecdn.net/M3uHhwSgLi9UkLi9R6qoH7.jpg https://www.gamesradar.com/black-mirror-bandersnatch-endings/ [2022. 10. 09.]

Jerusalem, Wilhelm 1916: *Bevezetés a filozófiába*. 2. kiadás. Pest: Franklin Társulat, Magyar Irod. Intézet és Könyvnyomda. 61–67.

Jia, Ye – Zhang, Yu – Weiss, Ron J. –Wang Quan – Shen Jonathan – Ren Fei – Chen Zhifeng – Nguyen Patrick – Pang, Ruoming – Lopez Moreno, Ignacio – Wu, Yonghui 2018: Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis. Google Inc. *Advances in Neural Information Processing Systems*, 31: 4485–4495. https://arxiv.org/abs/1806.04558 [2022. 10. 13.]

Keleti, Arthur 2016: *The Imperfect Secret*. Amazon, CreateSpace Independent Publishing Platform.

King, Thomas C. – Aggarwal, Nikita – Taddeo, Mariarosaria – Floridi, Luciano 2020: Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, 26: 89–120. https://doi.org/10.1007/s11948-018-00081-0 [2022. 10. 08.]

NIST (National Bureau of Standards) 1977: *Audit and Evaluation of Computer Security*. NBS Special Publication 500-19, U.S. Department of Commerce, National Bureau of Standards, október 26., 11–13. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nbsspecialpublication500-19.pdf [2022. 10. 05.]

Oord, Aaron van den – Dieleman, Sander – Zen, Heiga – Simonyan, Karen – Vinyals, Oriol – Graves, Alex – Kalchbrenner, Nal – Senior, Andrew – Kavukcuoglu, Koray 2016: WaveNet: A Generative Model for Raw Audio. Cornell University, Google DeepMind, London, UK, szeptember 12. *arXiv*:1609. 03499 [2022. 10. 10.]

Parker, Donn B. 1998: *Fighting Computer Crime*. New York, NY: John Wiley & Sons.

Pennycook, Gordon – Rand, David 2019: Why Do People Fall for Fake News? Are they blinded by their political passions? Or are they just intellectually lazy? *The New York Times*, január 19. https://www.nytimes.com/2019/01/19/opinion/sunday/fake-news.html [2022. 09. 30.]

Pew Research Center 2017: *The Fate of Online Trust in the Next Decade*. Pew Research Center, augusztus 10. www.pewresearch.org/internet/2017/08/10/the-fate-of-online-trust-in-the-next-decade/pi_17-08-10_onlinetrustnextdecade_0-01/ [2022. 10. 08.]

Ruggiero, Giuseppe – Zovato, Enrico – Di Caro, Luigi – Pollet, Vincent 2021: *Voice Cloning: a Multi-Speaker Text-to-Speech Synthesis Approach based on Transfer Learning*. Universita` degli Studi di Torino, Cerence Inc., 2021. február 10. https://arxiv.org/abs/2102.05630 [2022. 10. 12.]

Schilke, Oliver – Reimann, Martin – Cook, Karen S. 2021: Trust in Social Relations. *Annual Review of Sociology*, 240: 239–259.

Stanton, Brian – Jensen, Theodore 2020: *Trust and Artificial Intelligence*. NISTIR 8330, National Institute of Standards and Technology, U.S. Department of Commerce, december. https://doi.org/10.6028/NIST.XXXXXX [2022. 10. 07.]

Suwajanakorn, Supasorn – Seitz, Steven M. – Kemelmacher–Shlizerman, Ira 2017: Synthesizing Obama: Learning Lip Sync from Audio. *ACM Transactions on Graphics*, 36/4: 95.

Stupp, Catherine 2019: Fraudsters Use AI to Mimic CEO's Voice in Unusial Cybercrome Case. *The Wall Street Journal*, augusztus 30. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 [2022. 10. 10.]

Taraborelli, Dario 2008: How the Web Is Changing the Way We Trust. In: *Proceedings of the 2008 conference on Current Issues in Computing and Philosophy.* 194–204.

United States Air Force 1976: *Secure Computer System, Unified Exposition And Multics Interpretation.* Prepared for Deputy For Command And Management Systems, Electronic Systems Division, Air Force Systems Command, United States Air Force, Hanscom Air Force Base, Bedford, Massachusetts, March, 68, 70

Yao, Xinwei –Fried, Ohad – Fatahalian, Kayvon – Agrawala, Maneesh 2020: *Iterative Text-based Editing of Talking-heads Using Neural Retargeting.* Stanford University, The Interdisciplinary Center Herzliya. november 21. https://www.researchgate.net/publication/346143119_Iterative_Text-based_Editing_of_Talking-heads_Using_Neural_Retargeting [2022. 10. 11.]

Zakharov, Egor – Shysheya, Aliaksandra – Burkov, Egor – Lempitsky, Victor 2019: *Few-Shot Adversarial Learning of Realistic Neural Talking Head Models.* Cornell University, május 20. https://arxiv.org/abs/1905.08233v1 [2022. 09. 30.]

FORRÁSOK

Arik, Sercan O. – Chrzanowski, Mike – Coates, Adam – Diamos, Gregory – Gibiansky, Andrew – Kang, Yongguo – Li, Xian – Miller, John – Ng, Andrew – Raiman, Jonathan – Sengupta, Shubho – Shoeybi, Mohammad 2017: *Deep Voice: Real-time Neural Text-to-Speech.* Baidu Silicon Valley Artificial Intelligence Lab, Cornell University, február 25. https://arxiv.org/abs/1609. 03499 [2022. 09. 30.]

Brewster, Thomas 2021: Fraudsters Cloned Company Director's Voice In $35 Million Bank Heist, Police Find. *Forbes*, október 14. https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/ [2022. 10. 05.]

CTRL Shift Face, 2019: Bill Hader impersonates Arnold Schwarzenegger [DeepFake]. YouTube, május 11. https://youtu.be/bPhUhypV27w [2022. 10. 05.]

Peele, Jordan 2018: You Won't Believe What Obama Says In This Video! BuzzFeedVideo, YouTube, április 17. https://youtu.be/cQ54GDm1eL0 [2022. 10. 01.]

SZABADOS LEVENTE

Mélymerülés a „mélyhamisítás" világába

A deepfake-technológia jelene és (közel)jövője

Nem hihetünk a szemünknek. A „mélyhamisítás" (deepfake) technológiája, a 2012 óta tartó tudományos áttörés, a mély neurális hálós tanulás (deep learning) egy speciális ágaként 2017 óta tartó előretörésével az emberi percepciót megtéveszteni képes minőségben állít elő képeket, hangokat és mozgóképeket. A fejezet célja, hogy kvantitatív és kvalitatív eszközöket használva pontosabb képet alkosson a deepfake-technológia fejlődési dinamikájáról, társadalmi ismertségéről és alkalmazásáról, valamint becsléseket tegyen közeljövőbeli alakulásáról.

Kulcsszavak: technológia, fejlődési modellek, kvantitatív elemzés

1. BEVEZETÉS

A fejezet célja a deepfake mögötti technológiai kutatás, a fejlesztési ökoszisztéma és a deepfake téma szélesebb társadalmi ismertségének kvantitatív elemekkel alátámasztott elemzése, ezek alapján becslések megalkotása a technológia nagyjából hároméves távlatú közeljövőjére vonatkozóan.

1.1. Történeti kontextus

A deepfake-technológia mibenlétének jobb megragadásához elengedhetetlenül szükséges megértenünk a kialakulásának környezetét, hátterét, amelyet tágabb értelemben a mesterséges intelligencia kutatása, szűkebben pedig a gépi tanulás (azon belül is a gépi látás és hangfeldolgozás) szolgáltat.

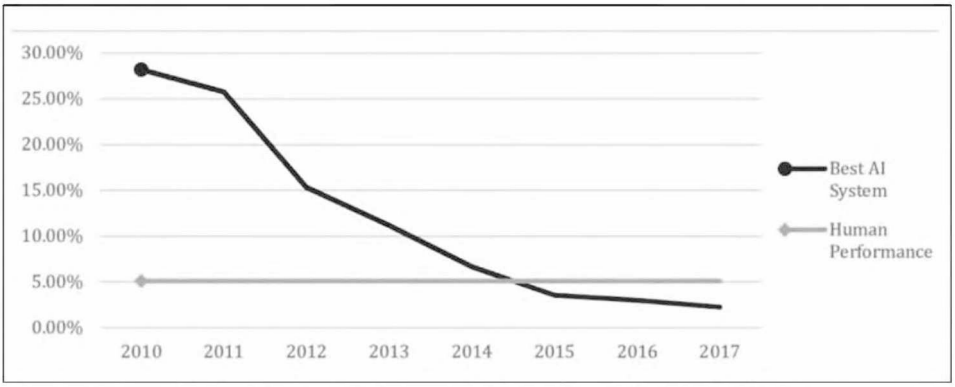
A számítástechnika történetében a legkorábbi számításorientált rendszereket szinte azonnal követték a perceptuális, általunk ismert érzékszervi feldolgozás területén tevékenykedő megoldások. Így például a gépi látás terén a Rosenblatt-féle Perceptron-modell (W1) már 1957-ben képes volt kezdetleges fényérzékelők segítségével közvetlenül a külvilágból származó képi bemenetek feldolgozására és osztályozására (Mason et al. 1958). A gépi látás tehát a mesterséges intelligencia

kutatása egyik legkorábbi területének is tekinthető. Érdekesség: e korai kísérletek kezdetleges neurális modellekkel történtek.

Ennek ellenére, némiképp a teljes mesterséges intelligencia terület időszakos „elszunnyadásával" („MI-tél"; Crevier 1993) párhuzamosan a gépi látás fejlődése nem volt töretlen. Hosszú ideig a szakemberek által manuálisan tervezett adatleírók („feature") és a rajtuk végzett statisztikai tanulási feladatok dominálták. E módszerek jellemzője, hogy egyrészről a teljesítmény szűk keresztmetszetét a befektetett szakértelem adja, azaz amennyi „mérnökórát" áldoznak az adott feladatra, azzal arányos, bár enyhe teljesítményjavulás várható. Másrészről az esetek jelentős részében „diszkriminatív" modellezési módszereket alkalmaztak, azaz főként vizuális (vagy akusztikus) bemenetek megkülönböztetését, osztályozását tűzték ki célul. A „generatív", vagyis vizuális (vagy akusztikus) jelek adott esetben realisztikus minőségben történő generálása, megalkotása jóval kevesebb hangsúlyt kapott. A fejlődés lassú volt és szakaszos, egészen a mélyhálós tanulás színre lépéséig.

Előbb az 1980-as évek kutatásai, főként Rumelhart, Hinton és Williams (1986) „hiba-visszaterjesztés" *(backpropagation of error)* algoritmusa, majd az 1990-es évek architekturális előrelépései, leginkább Le Cun (1998) „konvolúciós neurális hálói" *(convolutional neural net)* jelentős technikai előrelépéseket sejtettek. A paradigma lényege – a neurális modellekhez való visszatérésen túl – a szakértők által tervezett adatjellemzőktől vagy adatleíróktól *(feature)* való eltávolodás, azaz egy olyan rendszer ígérete, amely a többrétegű belső mesterséges neuronjai (tulajdonképpen számítási lépése vagy számítási gráf csomópontja) segítségével „mély" belső reprezentációt alakít ki a bejövő adatról, magából az adatból tanult módon transzformálja azt, így képes helyettesíteni a szakértői „manuális" munkát. A „mély tanulás" *(deep learning)* ígérete éppen ebben a belső, tanult reprezentációban rejlett, amelynek a kiaknázása főként az adatbázisok és a számítási kapacitás kezdetlegessége miatt egészen 2012-ig váratott magára.

2012-ben azonban az ILSVRC tudományos konferencián (W9) az azóta kvázi-szabványnak tekinthető ImageNet képfeldolgozási feladat (W3) a maga 20 000 osztályba sorolt 14 millió természetes képével már lehetővé tette, hogy Krizhevsky, Sutskever és Hinton (2012) alkalmasak legyenek meggyőzően demonstrálni nagy számítási teljesítmény bevetésével a mély neurális hálók dominanciáját ezen a felügyelt osztályozási, azaz diszkriminatív feladaton.



1. ábra. Hibaarány időbeli alakulása az ImageNet adatszetten (forrás: Kohler 2019)

E mélyreprezentáció sikere azonban komoly perspektívákat nyitott: újabb lendületet adott a generatív modellezés eddig némiképp mellőzött gondolatának is. Bár Le Cun már 1987-ben (Le Cun–Fogelman-Soulie 1987) felvetette a reprezentáció tanulásra fókuszáló, felügyeletlen „autoencoder" modellek lehetőségét, amelyek mindenféle emberi címkézés nélkül képesek tanulni a képek (és egyéb adatok) világáról úgy, hogy új, eddig nem látott képek létrehozására is alkalmasak legyenek, e terület a „denoising autoencoder" (Vincent et al. 2010) és a „variational autoencoder" (Kingma–Welling 2013) neurális modellek kidolgozásával indult fejlődésnek, végül Goodfellow (et al. 2014) munkássága révén vezetett el a „generatív ellenséges hálók" (generative adversarial networks, GAN) megalkotásához. Elmondható, hogy a GAN-modellek színre lépése egészen új szintre emelte – főként a képi területen – a generatív modellezést, vagyis az új, sosem látott képek egyre magasabb, napjainkban már fotorealisztikus minőségű megalkotásának a lehetőségét.

A mélyhálós GAN és utóbb a mélyhálós „diffúziós modellek" (*diffusion models, DDPM models*; Ho et al. 2020) képesek elképesztő minőségben felügyeletlen módszerrel megtanulni a képek mögött meghúzódó, azokat létrehozó szabályszerűségeket (statisztikai eloszlást), így a modern képgenerálás alapjait jelentik. Innen már csak egy lépés volt, hogy a generatív folyamat feletti kontroll megszerzésével és finomításával, nem egy „bármilyen", hanem egy „specifikusan valamilyen" kép létrehozásának képességével megszülessen a deepfake, azaz a mély neurális hálós „hamisítás" területe – és az ilyen generált képek felismerésének a gépi feladata.

1.2. Alapfogalmak és határterületek

A mélyhamisítás vagy deepfake módszerek pontos meghatározása első közelítésben csábítóan egyszerűnek tűnik: olyan mély neurális hálók segítségével alkotott média, amely nem a valóságot ábrázolja. Ez a munkadefiníció azonban több szempontból is elégtelen, hiszen nem vesz figyelembe több teljesen hétköznapi vagy teljességgel „ártatlan" szcenáriót, amely határos azzal, amit leginkább érdemes valóban deepfake-nek nevezni.

Az egyik oldalon a deepfake határterületét képezik azok a mély neurális hálós megoldások, amelyek szinte észrevétlenül meghúzódnak a mindennapi tevékenységeink mögött. Nyilván nem éljük meg „hamisítványnak" azt a képet, amelyet egy hétköznapi okostelefonnal készítünk, holott ha sötétebb környezetben használjuk, számos márka esetén már mély neurális hálók is dolgoznak a háttérben, hogy több képből egy, a valósnál sokkal világosabb és részletgazdagabb képet kapjunk, nem beszélve olyan felhasználási esetekről, amelyekben „nyuszifüleket" kapunk az önarcképre. Ebben az esetben ugyan a nyuszifül nem a mélyháló által generált képelem, ám az elhelyezéséhez modern mélyhálós arcfelismerő (diszkriminatív) algoritmusok szolgáltatták a koordinátákat. Ezen túlmenően azonban már megjelentek olyan megoldások, amelyekkel egy tájkép esetén egykönnyen „kicserélhetjük" az eget (Wankhede 2021) magán a telefonon, vagy akár tárgyakat tüntethetünk el generatív retusálással képszerkesztő programokban (W12), illetve mesterségesen megnövelhetjük egy korábban készített kép felbontását (W4). Ez utóbbi megoldás olyannyira elterjedt, hogy Deep Learning Supersampling (DLSS) néven az NVIDIA videokártyákban hardveresen beépült megoldás, azaz minden egyes képkocka egy modern játékban alacsony felbontáson „születik", és utána egy (generatív) mélyháló „gazdagítja fel" maximális felbontásra (W13). Bár sokkal kisebb számosságban, de audioterületen is lehetségesek felhasználási módok generatív mélytanulásra, mint amilyen a Deepmind Wavenet (van den Oord – Dieleman 2016) megoldása beszédhang létrehozására és „felolvasásra".

Annak ellenére, hogy a fenti esetekben a média része vagy akár minden egyes eleme is mélyhálókkal készült, ezeket a technológiákat nem soroljuk a deepfake körébe, bár technológiai szempontból is gyakorta tökéletesen megegyeznek annak módszereivel.

Javasolt munkadefiníciónk ezért a fenti naiv megközelítést két szempontból bővíti. Technikai aspektusból a deepfake elsődlegesen megfigyelhető és biometrikusan azonosítható, emberről szóló szignálok hamisítása, manipulációja – kizárva olyan nem közvetlenül az emberről szóló perceptuális domaineket vagy „másodlagos megfigyeléseket", mint az aláírás; társadalmi szempontból pedig a változtatás célja nem feltétlenül csupán egy tulajdonság megmásítása (mint fotóretusálásnál a modell hirtelen „fogyása"), hanem az illető ember identitásának (ottlétének, nemlétének vagy hollétének) vagy cselekedeteinek, közléseinek lénye-

gi megváltoztatása a szemlélő megtévesztésére. Ennek önkéntessége és konszenzusos mivolta (például az illető kérte-e és beleegyezett-e ebbe) már más kérdés. Egy „konszenzusos" deepfake is lehet deepfake.

2. MÓDSZERTAN

2.1. Kvantitatív alapok és kvalitatív értékelés

Más elemzésekkel (van Hooijdonk 2021) ellentétben megközelítésünk nem csupán kvalitatív módon kíván eljárni, hanem mérhető, számosságukban és időbeli alakulásukban informatív adatokkal kívánja alátámasztani megfigyeléseit. Bár a kvantitatív elemzések és a belőlük származó predikciók alapvető elemei a fejezetnek, mégsem mernénk állítani, hogy pontos becsléseket szeretnénk tenni egyes jelenségekről. A numerikus módszerek mankóként szolgálnak, ám interpretációjuk továbbra is elsődlegesen kvalitatív érvekre támaszkodik. A kvantitatív elemzések forráskódját nyilvánosan elérhetővé tettük (W19).

2.2. Adatforrások

Tudományos publikációk. Szerencsés fejlemény a gépi tanulás mint kutatási terület szempontjából, hogy fellendülése egybeesett a „nyílt tudomány" *(open science)* és a „nyílt forráskód" *(open source)* mozgalmak térnyerésével, így elmondható, hogy a kutatási publikációk szinte teljes mértékben nyilvánosak. E nyilvánosság legfőbb porondja az *ArXiv* pre-publikációs portál (W22). Ennek megfelelően a tudományos kutatás dinamikájának modellezéséhez az ArXiv keresőjéből származó adatokat nyertünk ki.

Open Source-aktivitás. Nevéből adódóan a nyílt forráskódú mozgalom – amely szerencsére a gépi tanulási terület domináns paradigmája is – megköveteli, hogy az egyének, közösségek, kutatócsoportok és cégek által fejlesztett forráskódot nyíltan elérhetővé tegyék. Ez esetünkben azt jelenti, hogy a deepfake területen felhasznált szoftvermegoldások és könyvtárak jelentős része is a „szabad szem előtt" fejlődik. E fejlődés legfőbb színtere a Github forrásmegosztó portál (W23), amely gyakorlatilag a modern informatikai fejlesztés gerince. Többek közt ezen a platformon érhető el a több milliárd példányban a telefonjainkon futó Android operációs rendszer forráskódja (Android Open Source Project, W10) vagy a világ informatikai infrastruktúráját és Android-telefonjait üzemeltető Linux (Linus Torvalds, W15) operációs rendszer is. A deepfake-technológia fejlesztési/fejlődési dinamikájának vizsgálatakor a Github keresőjére támaszkodtunk, és

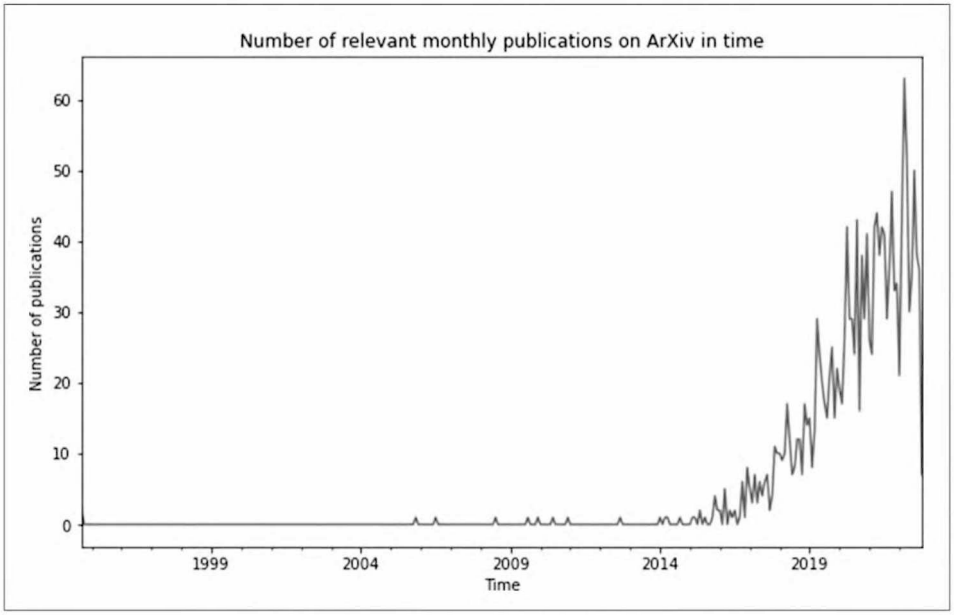
a „deepfake" keresőkifejezésre adott válaszként fellelhető „kódtárolók" *(repository)* fejlesztési aktivitását *(commit)* gyűjtöttük össze és használtuk fel az elemzéseinkhez.

Közismertség és közérdeklődés. Talán nem túlzás azt állítani, hogy napjainkban a közérdeklődés és a szélesebb közönség informálásának domináns médiuma az internet. Bár lehetetlen feladat lenne – már csak a különféle platformok üzemeltetőinek ellenérdekeltsége miatt is – egységes képet kapni arról, hogy deepfake témában a szélesebb közönség pontosan milyen információkhoz jut hozzá milyen médiumokból, ám azért plauzibilis azt állítani, hogy a téma iránti érdeklődés legalábbis erősen korrelál a publikum online keresési aktivitásával. Szerencsénkre a Google Trends szolgáltatás mind regionális, mind globális szinten összesíti és elérhetővé teszi a különféle keresőkifejezésekre történt keresések számszerű adatait. A közérdeklődés kvantifikálására tehát a „deepfake" keresőkifejezés globális gyakorisági adatait használtuk 2004-től napjainkig (Google Trends, W14). Bár a Google Trends a keresések pontos számát nem, csupán valamiféle arányszámot közöl, az adat mögött meghúzódó cselekvések száma mindenképp milliós nagyságrendű.

3. EREDMÉNYEK

3.1. A tudományos kutatás dinamikája

Ahogy azt jeleztük, megközelítésünk szerint az ArXiv preprint portálon történő pubikációk jól megragadják a tudományos kutatás dinamikáját. Felvethető, hogy nem minden tudományos publikáció követi a „nyílt tudomány" módszertanát, és jelenteti meg anyagait az ArXiv-on, sőt, azon is eltűnődhetünk, hogy az ArXiv népszerűsége az 1991-es alapítását követően fokozatosan nőtt, így akár ez irányú torzításokat is észlelhetünk, ám a képi (és hang)szintézisben megfigyelhető komoly minőségi változások időzítése alátámasztani látszik, hogy ez az adatsor jól tükrözi a kutatói érdeklődést.



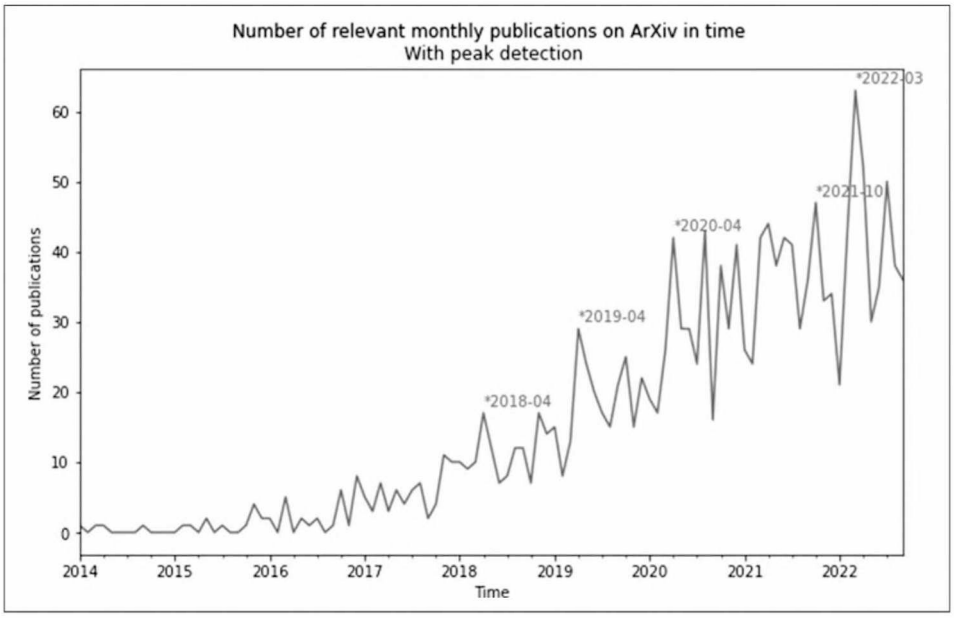
2. ábra. Releváns ArXiv-publikációk időbeli változása 1.
(forrás: saját szerkesztés)

Az erősen torzított ábrából kitűnik, hogy bár a képszintézissel foglalkozó legkorábbi publikáció már 1994-ben megjelent, a valódi érdeklődés 2014-től élénkült meg, a Goodfellow-féle GAN-modellek színre lépésével egy időben.

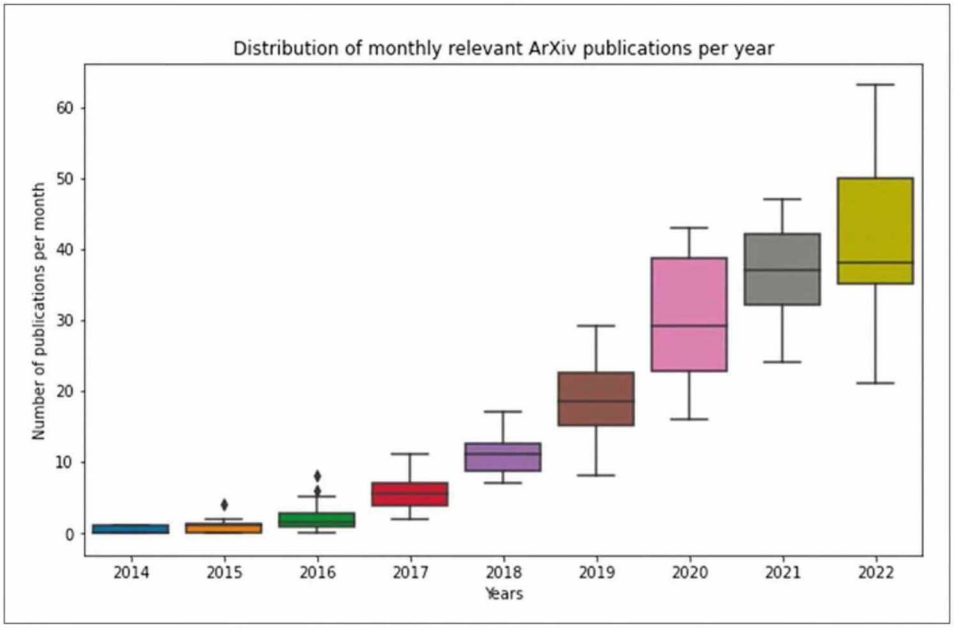
Közelebbről megvizsgálva az említett időszakot (és egy egyszerű csúcskeresési algoritmust lefuttatva, W18) láthatjuk, hogy a képgenerálás dominálta kép- és hangszintézis területén körülbelül évente történik komolyabb előrelépés, amelyet a publikációk számának hirtelen megugrása mutat. Míg a korábbi években feltehetően a GAN-modellek álltak az érdeklődés fókuszpontjában, az újabb csúcsok valószínűleg a diffúziós modellek színre lépésével magyarázhatók.

Emellett megállapítható – anélkül, hogy a lineáris, „fokozatos haladás" vagy az „exponenciális" kutatási aktivitás mellett elköteleznénk magunkat, bár utóbbira utalnak jelek, például az adatok szórásának növekedése az időben, multiplikatív trendet sugallva (González–Niet 2020) –, hogy a kutatói érdeklődés a téma iránt töretlen.

Fontos hangsúlyozni, hogy az adatsorba egyaránt belevettük általánosságban a deepfake modellek létrehozására alkalmas generatív technológiákat (mint a kép- vagy hangszintézis), specifikusan a deepfake-létrehozó modelleket, de a deepfake-detektálás automatikus modelljeit is. Ennek megfelelően nemcsak arról van szó, hogy egy feltehetőleg dinamikusan (akár exponenciálisan) növekvő generatív oldal, hanem egy ezzel párhuzamosan igen hangsúlyos diszkriminatív oldal is jelen



3. ábra. Releváns ArXiv-publikációk időbeli változása 2.
(forrás: saját szerkesztés)



4. ábra. Releváns ArXiv-publikációk időbeli változása 3.
(forrás: saját szerkesztés)

van, azaz afféle „fegyverkezési verseny" áll fenn a hamisításra alkalmas technikák és az ezek kiszűrésére alkalmas gépi technikák között. Ennek várható folyománya – hiszen a fotorealisztikus hamis képek valóditól való megkülönböztetése már jelenleg is egyre nehezebb (de még nem lehetetlen) az emberi szem számára –, hogy egyre komolyabban kell majd támaszkodnunk a gépi deepfake-detektálásra. Nem túlzás azt állítani, hogy kifejezetten rá leszünk szorulva olyan széles körben elérhető szolgáltatásokra, amelyek képesek a technológiai háttérrel nem rendelkező felhasználók – vagy legalább egy szűkebb körű, szakmailag elkötelezett szakértő, például médiamunkás vagy újságíró – számára megbízható jelet adni arról, hogy valós vagy „hamisított" anyagokkal áll-e szemben. Egyszerűbben fogalmazva: *meg kell szoknunk, hogy nem hihetünk a szemünknek, és a felismerésben digitális eszközökre kell támaszkodnunk.*

3.2. Az open source-fejlesztés dinamikája: eloszlás „madártávlatból"

Önmagában nézve az a tény, hogy a tudományos kutatás a deepfake téma szempontjából lendületesen halad, még nem jelentené azt, hogy a mögöttes technológia szélesebb körben elérhető és használható lenne. Lehetséges lenne, hogy a kutatási eredmények ugyan nyilvánosak, de azok reprodukálhatósága nehézkes – amit minden szakember, aki mások tudományos munkáját pusztán a publikációkból reprodukálni igyekezett, már a saját bőrén tapasztalhatta.

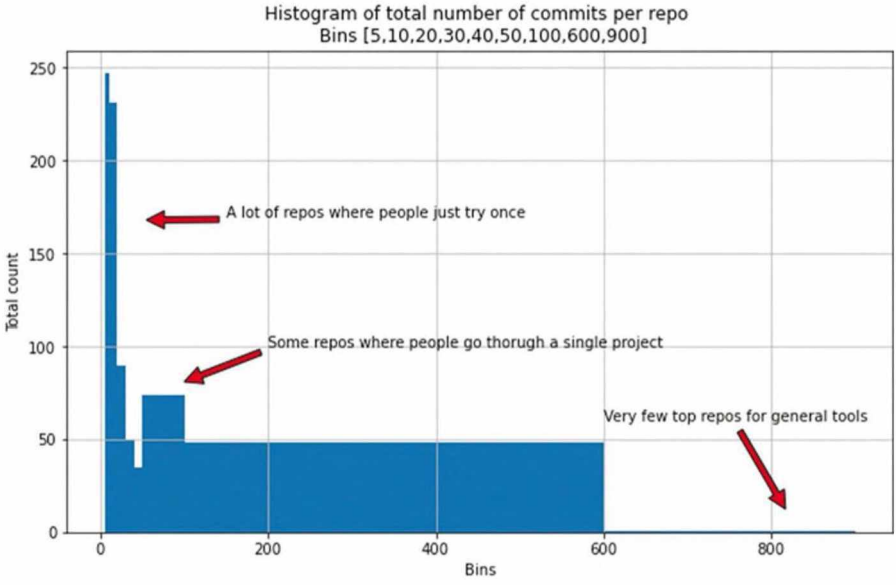
Állításunk szerint azonban nem ez a helyzet. Ha segítségül hívjuk a Github közösségi forráskódmegosztó portálról a kinyert adatokat, igen érdekes kép tárulhat elénk.

Az adat hátteréről érdemes megjegyezni, hogy a Github kódmegosztón minden egyéni felhasználó *(user)* képes létrehozni tetszőleges kódtárolót *(repository)*, amelyben ő és más felhasználók képesek együttműködni. Ennek az együttműködésnek az alapegysége az egységnyi „kódváltozás" vagy „hozzájárulás" *(commit)*. A keresés során lokalizált kódtárolókat elemeztük a „kódváltozások" összesített száma, illetve a kódtárolók létrehozási időpontjai szerint.

Ha csupán az összes kód „hozzájárulás" (commit) számát vizsgáljuk, sajátosan eltorzított eloszlást figyelhetünk meg. Bár összesen 1372 repository tűnik a keresés alapján a deepfake témába vágónak, igen élesen elkülöníthetünk három fő csoportot:

1. Az első, legnagyobb darabszámú csoportba tartoznak azok a repositoryk, amelyeket gazdáik csupán üresen vagy majdnem üresen, egy-egy apróbb kísérlet vagy programozási gyakorlás céljából hoztak létre, így gyakorlatilag jelentéktelenek, vagy legfeljebb enyhe érdeklődést szimbolizálnak.
2. A második csoport 50-100 committal egy-egy kísérlet vagy módszer reprodukcióját, kipróbálást és tanulást reprezentál.

	commit_count	first_commit_date
deepfakes/faceswap	1626	2017-12-15 11:47:32+00:00
samq-research/deepfakes-faceswap	1562	2017-12-15 11:47:32+00:00
Joannaever/deepfakes	1411	2017-12-15 11:47:32+00:00
iperov/DeepFaceLab	1309	2018-06-04 13:10:00+00:00
sgg423/deepfakes	1309	2018-06-04 13:10:00+00:00
ekohendratno/deepfakes-faceswap	1306	2017-12-15 11:47:32+00:00
wjqhqzmd/deepfake	1219	2017-12-15 11:47:32+00:00
big-bombom/deepfake	990	2018-06-04 13:10:00+00:00
aakashpadhiyar/deepfake-live	985	2017-12-15 11:47:32+00:00
yfsin/deepfake_pro	985	2018-06-04 13:10:00+00:00
Musa2077/DeepFaceLab	973	2018-06-04 13:10:00+00:00



5. ábra. A deepfake téma a Github felületén
(forrás: saját szerkesztés)

3. Néhány projekt 900 feletti commitszámmal kiemelkedik. Ezek az „érett", általánosabb eszköztár igényű projektek, amelyek számos módszert implementálnak, és törekszenek a szélesebb érdeklődők számára elérhetővé tenni. Ezekből összesen 11 darabot találtunk (a lista hirtelen eséssel 600 körüli commitszámokkal folytatódik).

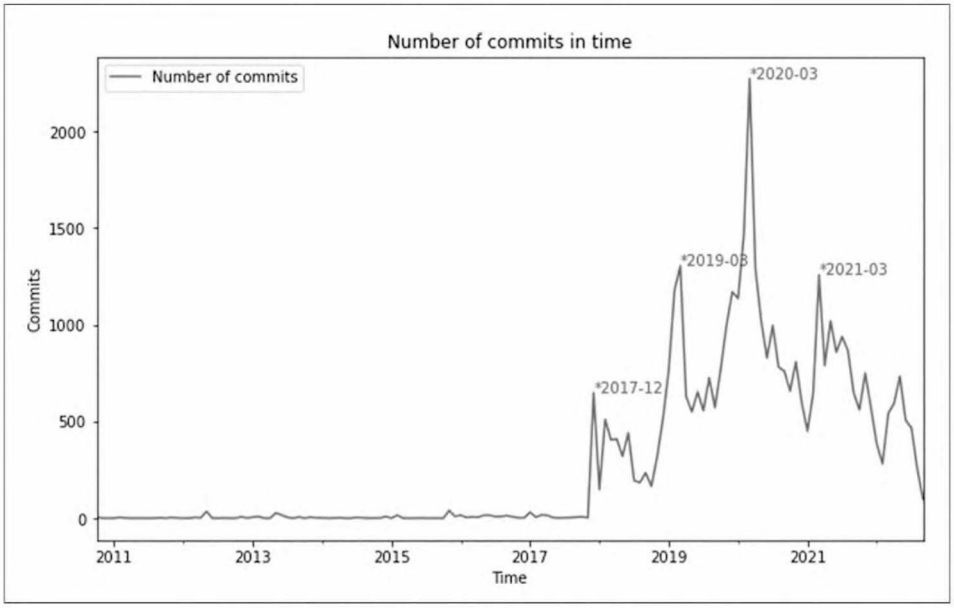
Feltehetjük, hogy ez az eloszlás egyfajta „a gazdag gazdagabbá válik" jellegű folyamat eredménye, azaz afféle érési folyamaté, amelyben az újonnan érkező érdeklődők lassanként egy-egy érettebb, általánosabb igényű kódkönyvtár köré csoportosulnak az idők során. (Megfigyelhető, hogy az összes „top" repo a 2017–2018-as időszakhoz kötődik, ekkor történt benne először bejegyzés.) Ez nyilván arra enged következtetni, hogy a deepfake-technológia kifejezetten érett, „konyhakész" megoldásokat nyújt, így a programozásban valamennyire jártas széles közeg is képes a „mélyhamisításra" – már ha befektetendő ideje, motivációja és legfőképp komputációs erőforrásokra fordítható büdzséje ezt megengedi. Egyszerűbben fogalmazva: *aki eléggé akarja, és van rá pénze és ideje, készíthet mélyhamisított tartalmakat.*

3.3. Az open source-közeg időbeli dinamikája

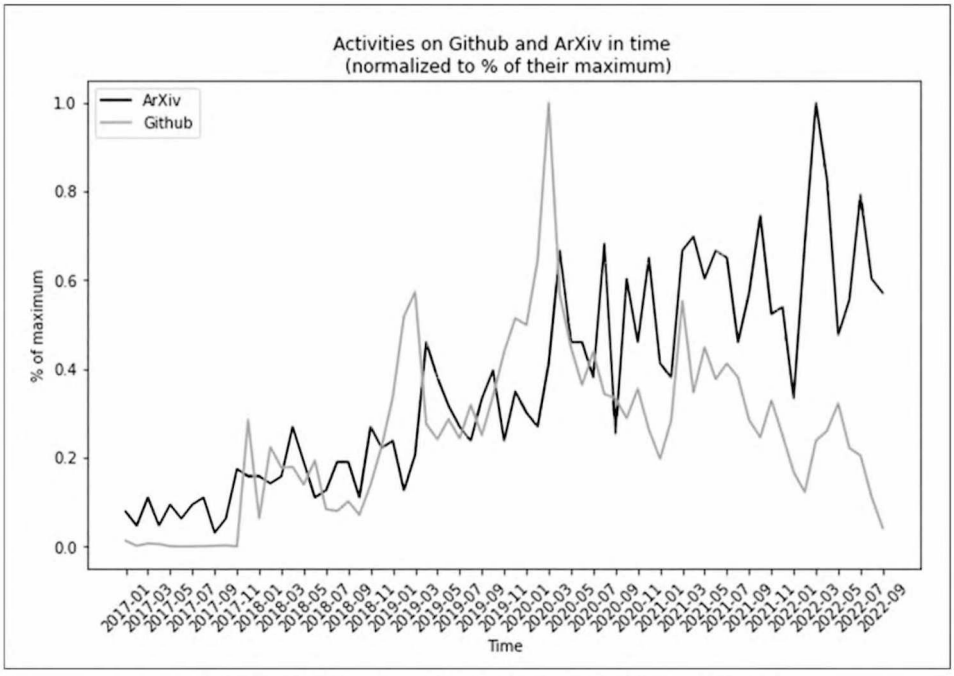
Az előző „felülről" való rátekintés mellett adja magát, hogy a nyílt forráskód közösség időbeli dinamikáját is megpróbáljuk áttekinteni, és megvizsgáljuk annak esetleges összefüggését a tudományos kutatás dinamikájával. Ehhez megvizsgáltuk a commitok, egyedi felhasználók és egyedi aktív repositoryk időbeli alakulását (aktív repositoryn olyan kódtárolót értünk, amelybe történt bejegyzés az adott hónapban).

Elsőként pusztán az összesített commitszámot vizsgálva néhány ismerős jelenségre figyelhetünk fel: hosszú időn át ez a téma szinte semmilyen mozgást nem mutatott, aztán 2017-ben egyszer csak berobbant, majd több hullámban mutat hirtelen fellángolásokat nagyságrendileg egyéves időritmusban. Ez gyanúsan hasonlít a tudományos kutatás korábbi dinamikájára.

Alaposabban szemügyre véve a két idősort, valóban van némi hasonlóság a szezonalitásukban, ritmusukban. Ahhoz, hogy ezt némiképp egzaktabban is vizsgálni tudjuk, segítségül hívhatjuk a Granger-féle idősoros tesztet (Granger causality test, W5), amely azt vizsgálja, hogy egy idősor mozgása jósló erővel bír-e egy másik idősor tekintetében, azaz valamiféle korlátozott értelemben oka-e a másik mozgásának. Maximum 12 hónapos interakciókat vizsgálva, mind a nem normalizált, mind a 7. ábrán is látható maximum százalékában normalizált adatokon a tesztek képesek kimutatni hatást, méghozzá a következőképpen: a Granger-teszt szignifikáns összefüggést jelez az ArXiv és a Github aktivitás közt, egy és hét hó-



6. ábra. A commitok számának változása
(forrás: saját szerkesztés)



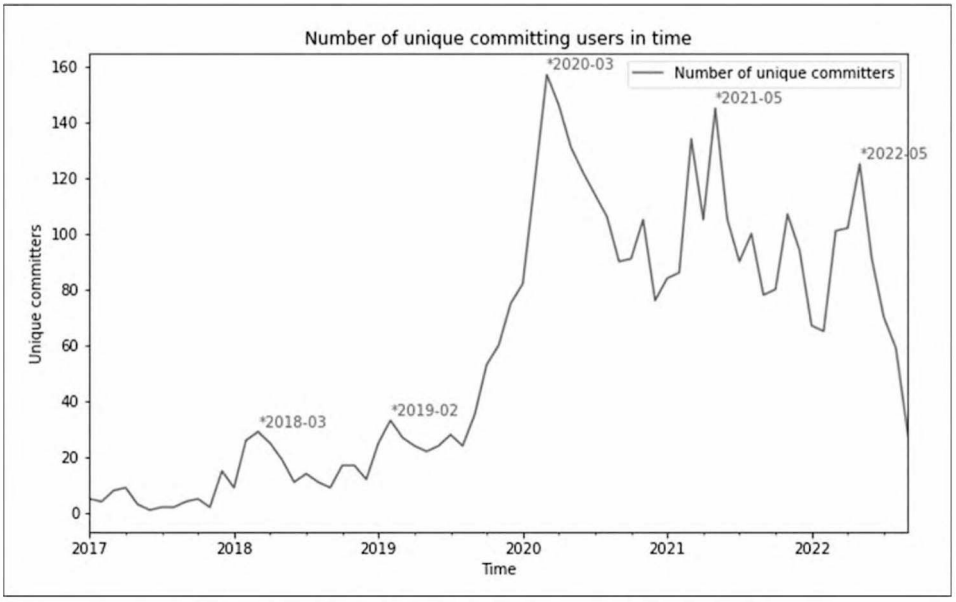
7. ábra. Az ArXiv- és a Github-aktivitások időbeli változása
(forrás: saját szerkesztés)

napos „késleltetéssel". A biztonsági tesztként alkalmazott Granger-vizsgálatok az ellenkező irányba (Github felől ArXiv felé) mind megbuktak.

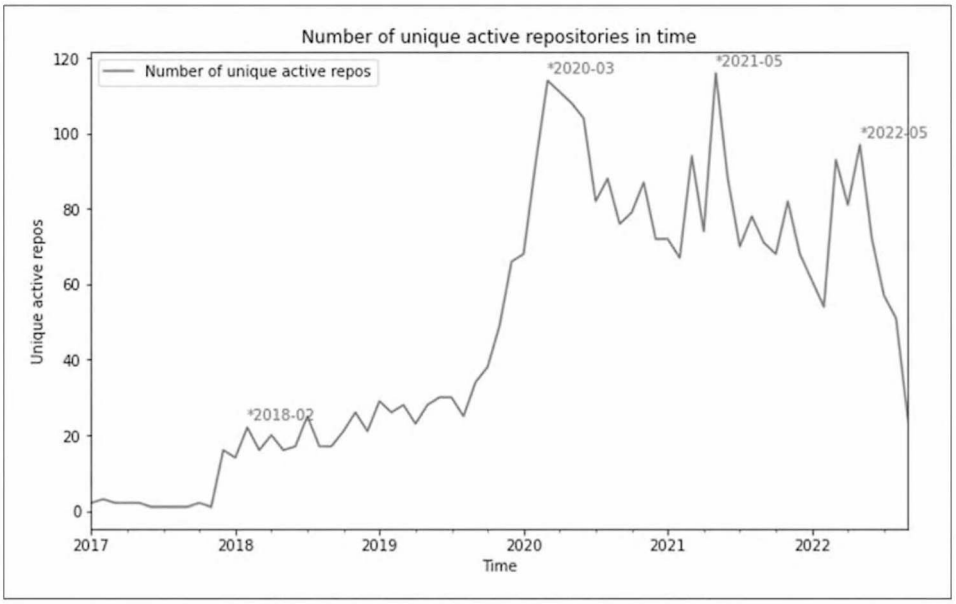
Mindezek fényében kijelenthető, hogy a tudományos kutatás előrelépései szinte azonnal, egy-hét hónap késéssel komoly hatással vannak az open source-közösségre, ami arra enged következtetni, hogy *„amit feltalálunk, az hamar közkézre kerül"*.

3.4. Divergencia: az open source-érdeklődés megváltozása

Az előző összefüggés kimutatása mellett azonban nem lehet nem észrevenni a Github-aktivitásban bekövetkezett trendszerű változást, vagyis azt, hogy a hozzájárulások száma a 2019-es csúcsához képest fokozatosan lefelé tartó trendet követ. Vajon milyen folyamatok állnak a háttérben? A tudományos publikációk száma kifejezetten nem csökken, így a szükséges „alapanyag" adott lenne. Nem valószínű, hogy a technológiai előrelépések hiánya (lásd ArXiv-idősor) lenne a fő akadály. Sokkal inkább lehetséges, hogy valamiféle professzionalizálódás áll a háttérben, azaz sok egyéni próbálkozás helyett egyre inkább kiemelkednek az átfogó keretrendszerek (vö. top 11 repository), amelyek immáron kevesebb egyéni kísérletezést kívánnak. Bár ahhoz, hogy ezt egyértelműen kijelenthessük, további vizsgálatokra lenne szükség, esetleges indikációként tekinthetünk arra, ha az összes commitszám mellett nem csak az aktív felhasználók, de az aktív repositoryk száma is csökken.



8. ábra. A Github-felhasználók számának időbeli változása
(forrás: saját szerkesztés)



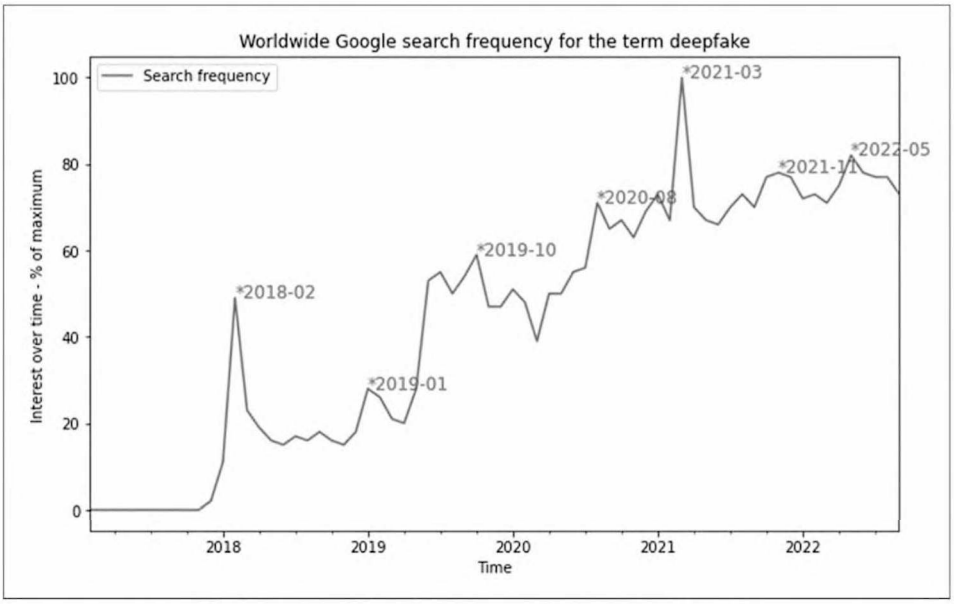
9. ábra. Az aktív repositoryk időbeli változása
(forrás: saját szerkesztés)

Amint azt a 8. és 9. ábrán is láthatjuk, valóban, az összaktivitással párhuzamosan az aktív felhasználók, de lényegesebben az aktív egyedi repositoryk száma is csökken. Bár egyértelmű bizonyítéknak ez kevés, de valamiféle centralizációt feltételezhetünk. Ha ezt elfogadjuk, akkor az lehet a jogos sejtésünk, hogy *egyre inkább „komoly", szofisztikált eszközök uralják a piacot.*

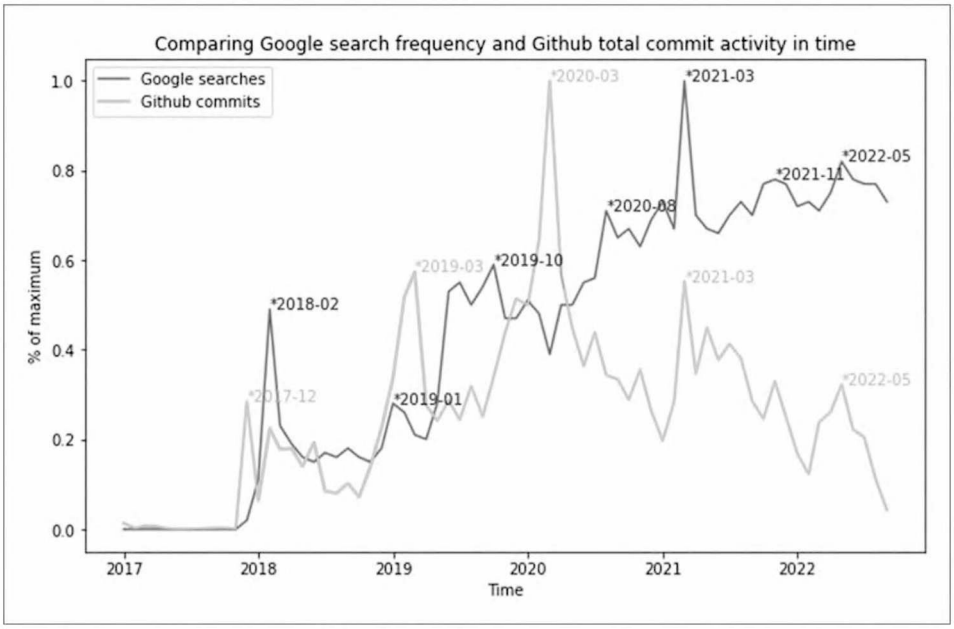
3.5. A közismertség dinamikája

Mi a helyzet azonban a „közismertséggel", azaz a széles körű, hangsúlyozottan nem szakmai nyilvánossággal, „az utca emberével"? A Google Trends keresési gyakoriságokat reprezentáló adatbázisa ebben segítségünkre lehet.

Az általános megállapítások a korábbiakhoz hasonlítanak: kis késéssel, 2018 februárjában a deepfake téma berobbant a köztudatba, majd folyamatos növekedésnek indul. Érdemes megvizsgálni az érdeklődés csúcsait, valamint összevetni őket az Open Source-aktivitással. Vajon a magasabb programozói aktivitás mozgatja-e a kereséseket, azaz az érdeklődést?



10. ábra. Google keresési gyakoriság a deepfake kifejezésre (forrás: saját szerkesztés)



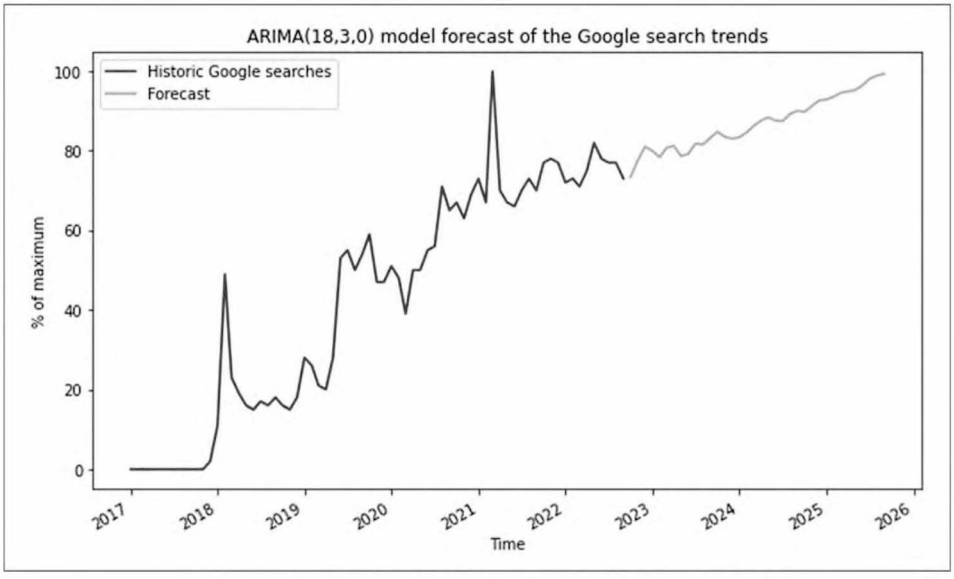
11. ábra. Google keresési gyakorisága és a Github commitok számának változása az időben (forrás: saját szerkesztés)

Megállapíthatjuk, hogy vannak átfedések, ám a kép sokkal vegyesebb a vártnál. Néhány kiugró csúcs igen közel esik egymáshoz, de mind késések, mind ritmusváltások is megfigyelhetők, nem beszélve a trendek fokozatos szétválásáról.

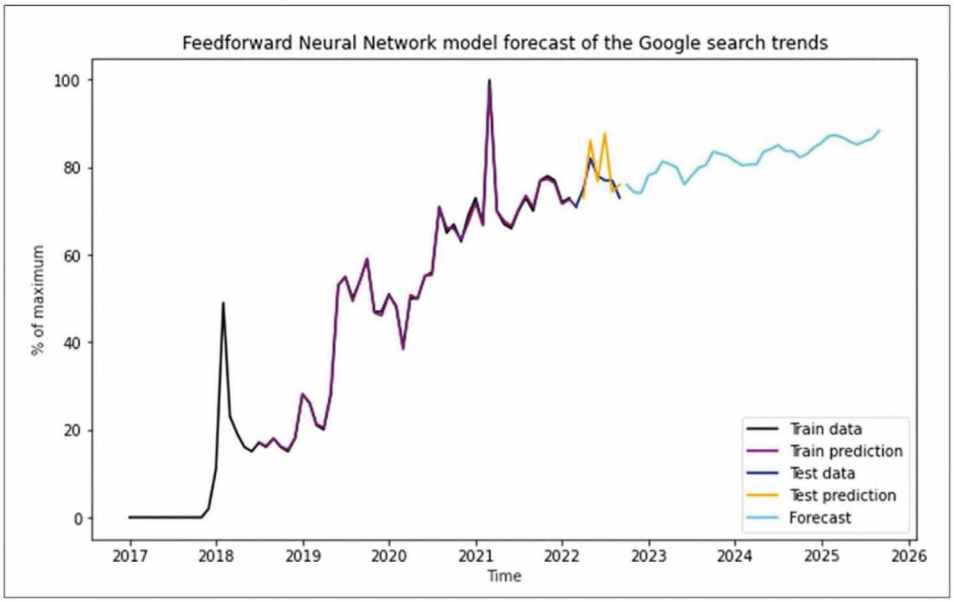
Formálisabban: megkísérelve a korábban is alkalmazott Granger-vizsgálatot a maximális 12 hónapos eltolási zónában, sem a Github-adatok felől a Google Trends felé, sem pedig fordítva nem tudunk szignifikáns és szisztematikus összefüggést kimutatni. Ebből arra következtethetünk, hogy talán vannak olyan események, hírek – és kifejezetten nem csak az új technológiák tudományos publikációjára gondolhatunk –, amelyek erős befolyást gyakorolnak a téma közismertségére, azaz: meglehet, *nem elég, hogy a technológia közkézen forog, kell olyan hír, valamilyen alkalmazásból adódó botrány, közéleti esemény, amely átlépi a széles nyilvánosság ingerküszöbét.*

3.6. Előrejelzés és hírelemzés

E lehetőséget két irányból is megvizsgáltuk. Egyrészről megkíséreltünk egyszerű idősoros modelleket illeszteni a Google Trends adatra, vakmerő módon hosszú, hároméves idősávra vetítve megkísérelni predikciókat tenni róla, modellezve ezzel a belső dinamikáját. Másrészről pedig megpróbáltunk felkutatni olyan híreseményeket, amelyek külső tényezőként, sokkhatásként megmagyarázhatják a kiugró csúcsokat és a közérdeklődés „lökéseit".



12. ábra. A Google-keresések és a prognózis ARIMA-modellel
(forrás: saját szerkesztés)



13. ábra. A Google-keresések és a prognózis neurális háló segítségével (forrás: saját szerkesztés)

Bár a modellezés és előrejelzés módszereinek részletes taglalásától most eltekintünk (W20), megállapíthatjuk, hogy az exponenciális trendet feltételező ARIMA-modell és az adatra illesztett mini neurális háló egyaránt emelkedő, némi szezonális ritmikával rendelkező mozgást vetít előre. Bár a modellek horizontja túlságosan hosszú, mégis annyit talán sejtethetnek, hogy „külső" lökések, a nagy nyilvánosság figyelmét megragadó hírek nélkül a deepfake téma iránti érdeklődés csupán fokozatosan élénkülne, nem rendelkezne a korábban megfigyelt erőteljes csúcsokkal. A jelek tehát arra mutatnak, hogy nem feltétlenül a technológiai innovációtól magától, de nem is a széles közönség fokozatos „edukálódásától" várhatjuk a legnagyobb hatást a közbeszéd formálásában, hanem sokkal inkább feltehető, hogy *egyes híreseményekadnak löketet, hogy egyre több és több ember figyeljen fel a deepfake létezésére.*

1. táblázat. Minta a deepfake-kel kapcsolatos angol nyelvű hírekből a vizsgált időszakban (forrás: saját szerkesztés)

Időszak	*Hírek*
2018-02	Reddit, Twitter Ban Deepfake Celebrity Porn Videos (Roetters 2018)
2019-01	Scarlett Johansson says fighting deepfake porn is ‚fruitless' (O'Brien 2019) Lawmakers warn of 'deepfake' videos ahead of 2020 election (O'Sullivan 2020)
2019-10	Meta AI (Facebook): „Creating a dataset and a challenge for deepfakes" (W16)
2020-08	MIT releases deepfake video of ‚Nixon' announcing NASA Apollo 11 disaster (Burton 2020) This is what a deepfake voice clone used in a failed fraud attempt sounds like (Vincent 2020)
2021-03	No, Tom Cruise isn't on TikTok. It's a deepfake (CNN, W11) Pennsylvania Woman Accused of Using Deepfake Technology to Harass Cheerleaders (Morales 2021)
2021-11	Adobe is toying around with deepfake tech for Photoshop (Paul 2021) Deepfake Audio Scores $35M in Corporate Heist (Lemos 2021)
2022-05	‚World first': Dutch police use ‚deepfake' video in appeal over boy's murder (Holroyd 2022) Deepfake Awareness Riding on Top Gun's Coattails (Robb 2022) Google bans deepfake-generating AI from Colab (Wiggers 2022)

Bár igen nehéz formálisan is bizonyítani – így mi sem állítjuk –, hogy ezek, éppen ezek és csak ezek a hírek (1. táblázat) okozták volna a közérdeklődés megugrását, annyit mindenesetre jól illusztrálnak, hogy főként a közéleti személyek támadása, a csalási események, illetve a technológiai vállalatok deepfake-re adott reakciói komoly sajtóvisszhangra találnak, így alkalmasak újabb „lökéseket" adni a téma közismertségi folyamatának. Becslésünk, hogy ilyen események nélkül az érdeklődés növekedése csupán fokozatos lenne, míg a híresemények hirtelen „szintlépéseket" okoznak.

4. KÖVETKEZTETÉSEK ÉS BECSLÉSEK

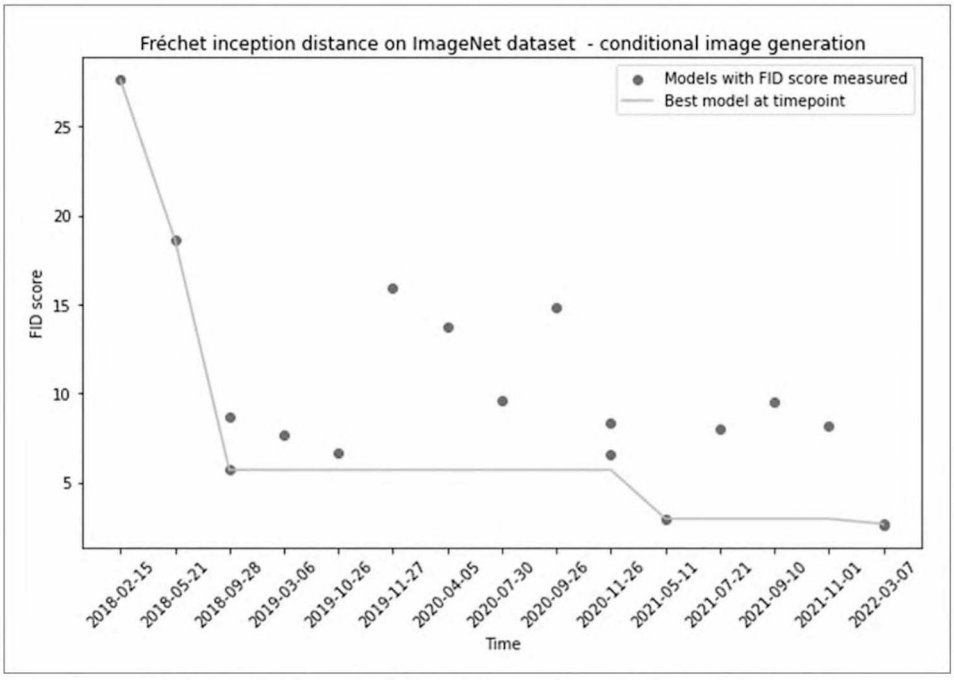
A tudományos publikációk, az Open Source-aktivitás és a közérdeklődés dinamikájának leírása után ideje kvalitatív becsléseket is megkísérelnünk, mintegy „megtéve tétjeinket" a közeljövővel kapcsolatban.

4.1. A technológiai fejlődés közeljövője

A deepfake-technikák alapja a jó minőségű generatív modellek megléte. Érdekes módon azonban a generált képek minőségének megállapítása nem olyan triviális feladat, mint egy diszkriminatív modell (és az ehhez kapcsolódó címkézett adat-

bázis) esetén, hiszen míg egy osztályozási feladatnál rendelkezésünkre áll valamiféle emberileg rögzített címke („ez a kép egy macskát ábrázol, az meg egy kutyát"), amely minden esetleges hibája ellenére kevéssé vita tárgya, addig az újonnan létrehozott képek „realisztikusságát" nehéz még tanult címkéző szakembereknek is egységes minőségben és objektív módon végezni. Ennek megfelelően – és némiképp a deepfake területen megfigyelhető „macska-egér játékkal" párhuzamban – a kutatók úgy döntöttek, hogy előre betanított képosztályozó modelleket használnak a generált képek minőségének mérésére. Egyszerűbben fogalmazva: ha egy képosztályozó modell elfogadja, hogy egy macska van a generált képen, akkor felvethető, hogy a generált macskakép realisztikus. Éppen ezért a jelenlegi kutatások terén a generatív modellek jósága egy fixen a standard ImageNet adatbázison betanított Inception nevű képosztályozó modell (Szegedy et al. 2014) elfogadási aránya *(inception score)* vagy belső reprezentációs terében végzett valós és generált képek közötti távolság mérése (Fréchet Inception Distance, FID; Heusel 2018) alapján ítélhető meg.

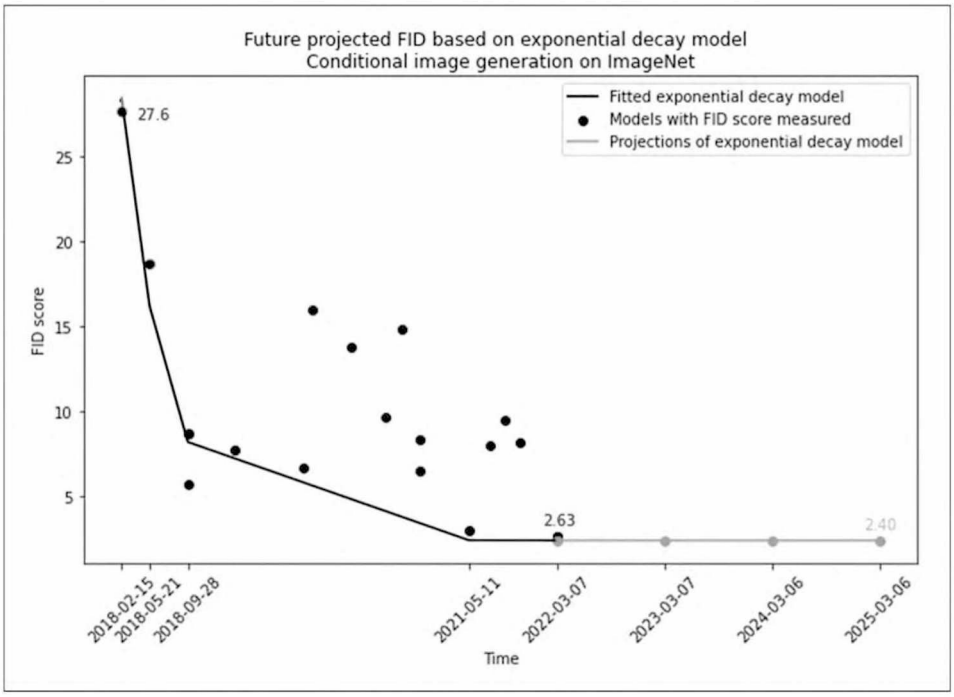
A közelebbi jövő (a következő három év) tudományos előrelépéseinek becslésére ezért a Papers with Code nevű tudományos gyűjtőoldalon közzétett (W6), az ImageNet adatbázison végzett, osztályalapon kondicionált (értsd: „hozz létre új sziámimacska-képet") képgenerálási feladat FID-értékeit használjuk és elemezzük.



14. ábra. Fréchet kezdeti távolság (FID) az ImageNet adatkészleten (forrás: saját szerkesztés)

A 14. ábrán a FID-metrika időbeli alakulását láthatjuk, amiből kitűnik, hogy bár az újabb és újabb tudományos publikációk (pontok) nem mindig csak a FID-metrika (azaz a „perceptuális jóság") szempontjából innovatívak, hanem más, például számítási teljesítmény szempontjából, de még így is néhány évente radikális előrelépések látszanak, azaz tulajdonképpen egy exponenciálisan javuló folyamatot láthatunk (emlékeztetőül: a FID-metrika esetén a „minél alacsonyabb, annál jobb" összefüggés érvényesül, mivel annál közelebb esik a generált kép az ImageNet valóságához).

Ezen felbátorodva egy igen egyszerű exponenciálisgörbe-illesztést alapul véve az alábbi eredményekhez juthatunk (a modellezés részletei: W21):



15. ábra. Az FID pontszám várható alakulása az időben az ImageNet képgenerálási feladaton (exponenciális lecsengési modell alapján, forrás: saját szerkesztés)

Hároméves becslésünk egy 2,40-es FID-pontszám, amely lényegében nagyon hasonló a jelenlegi legjobb, 2,63-as értékhez.

Ebből két fő következtetés vonható le: Egyrészről elmondható, hogy a fejlődés a korábbi drámai (exponenciális) minőségi ugrást követően „szaturálódott", azaz már ugyanilyen drasztikus előrelépések nem várhatóak. Ez azt jelenti, hogy már így is *megdöbbentően közel vagyunk a fotorealisztikus szinthez, amely em-*

16. ábra. Egy diffúziós modell által, szabad szöveges utasításra generált kép (forrás: Strikingloo 2022)

pedig a technológia kerete adott, várható, hogy nemsokára széles körben deepfake témában is hasonlóan járhassunk el, és ne csak kedves rajzfilmfigurákat legyünk képesek generálni. Kérdés, ezt szeretnénk-e, és ha nem, mit tehetünk…

4.2. A társadalmi beágyazottság közeljövője

A technológia közkézre kerülése és leegyszerűsödése ellen – állításom szerint – nem sokat. Az itt bemutatott elemzések jól illusztrálták, hogy milyen folyamatok során és mekkora sebességgel kerülnek át a főbb innovációk a gyakorlatba, azaz egyre szélesebb körben használható formába. Ebből következik, hogy a deepfake – vagy nevezzük inkább algoritmikusan támogatott kreativitásnak – újabb formáival együtt kell élnünk, a társadalomnak ki kell alakítani a normák és szabályok rendszerét, amely éppúgy, mint minden más technológia terén, kijelöli a legális és társadalmilag üdvös felhasználások körét, valamint módszereket fejleszt, tesz el-

érhetővé vagy egyenesen kötelezővé a károk megakadályozására vagy az esetleges normasértések (netán bűncselekmények) szankcionálására.

Anélkül, hogy a szükséges változások teljes áttekintését kísérelnénk meg, néhány pontot érdemes kiemelnünk:

- A szellemet nem lehet már visszatömni a palackba, mivel széles körben elterjedt technológiáról van szó.
- Nemsokára valóban nem hihetünk a szemünknek, mert elterjedt és minőségi lesz a deepfake.
- Támaszkodnunk kell azonosító szolgáltatásokra, amelyek gépileg validálják, hogy amit látunk, valós-e.
- Valamiféle jogi, viselkedéses és politikai konszenzusnak kell kialakulnia, amivel szabályozni tudjuk a deepfake okozta társadalmi hatásokat.
- Folyamatos fejlődés várható a közérdeklődésben, de főként események, esetleg sokkok kellenek a szélesebb körű érdeklődés felkeltéséhez. Ilyen események megjelenhetnek akár a pozitív oldalon is. Például a 2022. őszi hír szerint Bruce Willis a betegségére tekintettel átadta egy technológiai startupnak a képmásához fűződő jogait, hogy az deepfake formájában szerepeltethesse őt későbbi filmekben (Dent 2022, W24). Bár a hír futótűzként terjedt el, nem sokkal később cáfolták (Quach 2022). Mindenesetre tekinthető újabb „lökésnek" a társadalmi ismertségben.
- Fokozatosan a társadalmi tudatosságban széles körben elterjed majd a deepfake létezésének a tudata és lehetőségei.

Ideje készülni. Semmi nem az, aminek látszik.

SZAKIRODALOM

Crevier, Daniel 1993: *AI: The Tumultuous Search for Artificial Intelligence*. New York, NY: BasicBooks.

Krizhevsky, Alex – Sutskever, Ilya – Hinton, Geoffrey E. 2012: ImageNet classification with deep convolutional neural networks. *NIPS 12: Proceedings of the 25th International Conference on Neural Information Processing Systems*. Volume 1. 1097–1105. https://dl.acm.org/doi/10.5555/2999134.2999257

Le Cun, Yann – Fogelman-Soulie, Françoise 1987: Modèles connexionnistes de l'apprentissage Intellectica. *Année*, 2–3: 114–143. https://www.persee.fr/doc/intel_0769-4113_1987_num_2_1_1804

Rumelhart, David E. – Hinton, Geoffrey E. – Williams, Ronald J. 1986: Learning representations by back-propagating errors. *Nature*, 323: 533–536. https://www.nature.com/articles/323533a0; https://doi.org/10.1038/323533a0

Szegedy, Christian – Liu, Wei – Jia, Yangqing – Sermanet, Pierre – Reed, Scott – Anguelov, Dragomir – Erhan, Dumitru – Vanhoucke, Vincent – Rabinovich, Andrew 2014:

Going Deeper with Convolutions. *ArXiv.* https://arxiv.org/abs/1409.4842, https://doi.org/10.48550/arXiv.1409.4842

Vincent, Pascal – Larochelle, Hugo – Lajoie, Isabelle – Bengio, Yoshua – Manzagol, Pierre-Antoine 2010: Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research*, 11: 3371–3408. https://www.jmlr.org/papers/volume11/vincent10a/vincent10a.pdf

FORRÁSOK

(minden forrás utolsó ellenőrzési dátuma: 2022. 10. 24.)

Abram, Cleo 2022: The REAL fight over AI art. https://www.youtube.com/watch?v=NiJeB2NJy1A

Burton, Bonnie 2020: MIT releases deepfake video of ‚Nixon' announcing NASA Apollo 11 disaster. CNET, július 20. https://www.cnet.com/science/mit-releases-deepfake-video-of-nixon-announcing-nasa-apollo-11-disaster/

Dent, S. 2022: A Bruce Willis deepfake could appear in his stead for future film projects (updated). https://www.engadget.com/bruce-willis-deepfake-celebrity-rights-192200856.html

González, Joaquín – Niet, Fabio H. 2020: Bayesian Analysis of Multiplicative Seasonal Threshold Autoregressive Processes. http://www.scielo.org.co/pdf/rce/v43n2/0120-1751-rce-43-02-251.pdf

Goodfellow, Ian J. – Pouget-Abadie, Jean – Mirza, Mehdi – Xu, Bing – Warde-Farley, David – Ozair, Sherjil – Courville, Aaron – Bengio, Yoshua 2014: Generative Adversarial Networks. https://arxiv.org/abs/1406.2661

Heusel, Martin 2018: GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. https://arxiv.org/abs/1706.08500

Ho, Jonathan – Jain, Ajay – Abbeel, Pieter 2020: Denoising Diffusion Probabilistic Models. https://arxiv.org/abs/2006.11239

Holroyd, Matthew 2022: ‚World first': Dutch police use ‚deepfake' video in appeal over boy's murder. https://www.euronews.com/my-europe/2022/05/23/world-first-dutch-police-use-deepfake-video-in-appeal-over-boy-s-murder

Kingma, Diederik P. – Welling, Max 2013: Auto-Encoding Variational Bayes. https://arxiv.org/abs/1312.6114

Kohler, Kevin 2019: The Construction of Artificial Intelligence in the U.S. Political Expert Discourse. Master's Thesis. Master of Arts in International Affairs and GovernanceUniversity of St. Gallen. https://www.researchgate.net/publication/340502861_The_Construction_of_Artificial_Intelligence_in_the_US_Political_Expert_Discourse/figures?lo=1

Le Cun, Yann 1998: http://vision.stanford.edu/cs598_spring07/papers/Lecun98.pdf

Lemos, Robert 2021: Deepfake Audio Scores $35M in Corporate Heist. https://www.darkreading.com/attacks-breaches/deepfake-audio-scores-35-million-in-corporate-heist

Mason, Harding – Stewart, D. – Gill, Brendan 1958: Rival. *The New Yorker*, december 6. 44. https://www.newyorker.com/magazine/1958/12/06/rival-2

Morales, Christina 2021: Pennsylvania Woman Accused of Using Deepfake Technology to Harass Cheerleaders. *The New York Times*, március 14. https://www.nytimes.com/2021/03/14/us/raffaela-spone-victory-vipers-deepfake.html

O'Brien, T. 2019: Scarlett Johansson says fighting deepfake porn is ‚fruitless'. https://www.engadget.com/2019-01-01-scarlett-johansson-fighting-deepfake-porn-lost-cause.html

O'Sullivan, Donie 2020: Lawmakers warn of 'deepfake' videos ahead of 2020 election. CNN, 2019. január 28. https://edition.cnn.com/2019/01/28/tech/deepfake-lawmakers/index.html

Paul, Andrew 2021: Adobe is toying around with deepfake tech for Photoshop. https://www.inputmag.com/tech/adobe-is-toying-around-with-deepfake-tech-for-photoshop

Perktold, Josef – 2019: Statsmodels https://www.statsmodels.org/stable/index.html

Quach, Katyanna 2022: This rumor needs to Die Hard: Bruce Willis denies selling face to deepfake biz. https://www.theregister.com/2022/10/04/bruce_willis_ai_image_deepcake/

Robb, Drew 2022: Deepfake Awareness Riding on Top Gun's Coattails. https://www.cioinsight.com/news-trends/deepfake-awareness-top-gun/

Roetters, Janko 2018: Reddit, Twitter Ban Deepfake Celebrity Porn Videos. Variety, https://variety.com/2018/digital/news/reddit-twitter-deepfake-ban-1202690627/

Van den Oord, Aäron – Dieleman, Sander 2016: WaveNet: A generative model for raw audio. Deepmind, szeptember 8. https://www.deepmind.com/blog/wavenet-a-generative-model-for-raw-audio

Van Hooijdonk, Richard 2021: https://blog.richardvanhooijdonk.com/en/the-good-the-bad-and-the-future-of-deepfakes/

Vincent, James 2020: This is what a deepfake voice clone used in a failed fraud attempt sounds like. *The Verge*, július 27. https://www.theverge.com/2020/7/27/21339898/deepfake-audio-voice-clone-scam-attempt-nisos

Vincent, James 2022: Anyone can use this AI art generator – that's the risk. The Verge, szeptember 15. https://www.theverge.com/2022/9/15/23340673/ai-image-generation-stable-diffusion-explained-ethics-copyright-data

W1 = https://psycnet.apa.org/record/1959-09865-001

W3 = Deng, J. – Dong, W. R. – Socher, L. -J. – Li, Kai Li – Li Fei-Fei 2009: ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA. 248–255. DOI: 10.1109/CVPR.2009.5206848

W4 = Behind the Tech: Super Resolution in Adobe Photoshop and Lightroom 2022: https://research.adobe.com/news/behind-the-tech-super-resolution-in-adobe-photoshop-and-lightroom/

W5 = Granger causality https://en.wikipedia.org/wiki/Granger_causality

W6 = Conditional Image Generation on ImageNet 128x128 https://paperswithcode.com/sota/conditional-image-generation-on-imagenet

W7 = Stable Diffusion Public Release https://stability.ai/blog/stable-diffusion-public-release

W8 = Dickey–Fuller test. https://en.wikipedia.org/wiki/Dickey%E2%80%93Fuller_test

W9 = Stanford University 2012: ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012). https://www.image-net.org/challenges/LSVRC/2012/

W10 = Android Open Source Project: https://github.com/aosp-mirror

W11 = CNN 2021: No, Tom Cruise isn't on TikTok. It's a deepfake. CNN, március 2. https://edition.cnn.com/videos/business/2021/03/02/tom-cruise-tiktok-deepfake-orig.cnn-business
W12 = sz. n. 2021: Blog: Nvidia Image Inpainting. Digital Meat, YouTube. https://www.youtube.com/watch?v=poJx7vK3mLo
W13 = DLSS, Nvidia 2022: https://www.nvidia.com/nl-nl/geforce/technologies/dlss/
W14 = Google Trends: https://trends.google.com/trends/explore?date=all&q=deepfake [2022.10.24.
W15 = Linus Torvalds: https://github.com/torvalds/linux
W16 = MetaAI 2019: Creating a dataset and a challenge for deepfakes https://ai.facebook.com/blog/deepfake-detection-challenge/
W17 = Strikingloo 2022: Stable Diffusion: Prompt Guide and Examples. https://strikingloo.github.io/stable-diffusion-vs-dalle-2
W18 = Szabados Levente 2022: https://github.com/solalatus/deepfake_popularity/blob/main/arxiv_analyzer.ipynb
W19 = Szabados Levente 2022: A kvantitatív elemzések forráskódja. https://github.com/solalatus/deepfake_popularity
W20 = Szabados Levente 2022: https://github.com/solalatus/deepfake_popularity/blob/main/global_analysis.ipynb
W21 = Szabados Levente 2022: https://github.com/solalatus/deepfake_popularity/blob/main/paperswithcode_downloader.ipynb
W22 = https://arxiv.org/
W23 = https://github.com/
W24 = https://tilda.cc/
Wankhede, Calvin 2021: How on-device machine learning has changed the way we use our phones. Android Authority, július 23. https://www.androidauthority.com/machine-learning-phones-1238287/
Wiggers, Kyle 2022: Google bans deepfake-generating AI from Colab. https://techcrunch.com/2022/06/01/2328459/

KRASZNAY CSABA

A deepfake-technológia kiberbiztonsági vonzatai

A mesterséges intelligencia témakörébe tartozó technológiák évek óta komoly kihívás elé állítják a kiberbiztonsági szakembereket. Ezen belül a deepfake az egyik olyan megvalósítás, amelynek szerepéről és veszélyeiről számos tanulmány született, de érdemi technológiai védelem még nem terjedt el széles körben. Eközben a kiberbűnözés, a kiberhadviselés, a kiberkémkedés és a hacktivizmus területén sorra látnak napvilágot olyan hírek, amelyek arra utalnak, hogy a támadó oldal aktívan használja ezt a megoldást. Különös jelentőséget ad a témaválasztásnak az orosz–ukrán háború, amelyben az összecsapások elején mindkét fél élt a deepfake-kel. A fejezet célja esettanulmányokon keresztül áttekinteni a deepfake valódi kockázatait 2022–2023-ban és rámutatni, milyen megoldásokkal lehet a kibertér védelmét fejleszteni e specifikus kockázattal szemben.

Kulcsszavak: kiberhadviselés, kiberbűnözés, kiberkémkedés, hacktivizmus

1. KIBERBIZTONSÁGI KIHÍVÁSOK 2022-BEN

A deepfake-technológiát számos szempontból lehet megvizsgálni, de hátulütői közül valószínűleg a legtöbb elemzés elsőként a kiberbiztonsági kihívásokat fogja említeni. Holott kiberbiztonsági szempontból a deepfake csak egy eszköz, amelyet viszont széles körben lehet alkalmazni többfajta támadásnál. A kiberbiztonság alapvetése, hogy a legtöbb problémát „a szék és a billentyűzet között kell keresni", azaz az emberi hibák nélkül nehezen lenne elképzelhető egy sikeres, kibertérből érkező csapás. Jellemzően ugyanis a támadónak arról kell meggyőznie egy célszemélyt, hogy az nyisson meg egy kártékony kódot tartalmazó üzenetet, vagy adjon át egy információt, amelynek segítségével be lehet jutni a kiszemelt hálózatba. Az embert támadó módszereket összefoglaló néven *social engineering*nek vagy emberi ráhatással történő támadásnak, esetleg pszichológiai manipulációnak nevezik, és ebben a deepfake-technológia kiváló eszköznek bizonyulhat a támadó kezében.

A social engineering lényege, hogy valamilyen emberi tulajdonságot kihasználva jut a támadó az őt érdeklő információhoz. Kevin Mitnick, az egyik legis-

mertebb social engineer szavaival: „a social engineering a befolyásolás és a rábeszélés eszközével megtéveszti az embereket, manipulálja, vagy meggyőzi őket, hogy a social engineer tényleg az, akinek mondja magát. Ennek eredményeként a social engineer – technológia használatával vagy anélkül – képes az embereket információszerzés érdekében kihasználni" (Mitnick 2003). Oroszi Eszter két típusra bontotta ezeket a technikákat: humán- és számítógép-alapú támadásokra (Oroszi 2019: 87). A deepfake mindkét esetben hatékony lehet, hiszen például a megtévesztés/megszemélyesítés vagy az álweboldalak létrehozása esetén is jól használható a deepfake. A vízválasztó az, hogy mennyire lehet valós időben alkalmazni ezt a mesterséges intelligencia által generált alternatív valóságot. A nem valós időben létrehozott deepfake-ek is igen sikeresek, de az igazán hatékony social engineering támadás megköveteli az azonnaliságot, például egy telefonbeszélgetés vagy egy videócset használata során.

Az információszerzés, illetve a klasszikus kibertámadások mellett azonban a social engineering jól használható egyes katonai műveletek során is, különösen a lélektani műveletek *(Psychological Operations, PSYOPS)* területén. Ahogy Deák Veronika fogalmaz, a „lélektani műveletek alapvető célja a másik fél befolyásolása, amely eléréséhez számtalan, a kibertér felhasználásán alapuló módszer vehető igénybe, egyik jellemző formájuk az emberi tényező és az infokommunikációs eszközök gyengeségeit, illetve sérülékenységeit együttesen kihasználó támadási módszer, a social engineering (a továbbiakban: SE). A social engineering módszerek alkalmazásával jelentősen növelhető a lélektani műveletek kivitelezéséhez szükséges információk megszerzésének mennyisége és minősége, továbbá ezen műveletek során végrehajtott befolyásolás minősége és hatékonysága" (Deák 2019: 97). Nem véletlen tehát, hogy a napjainkban zajló orosz–ukrán háború során mindkét fél és az őket támogató harmadik felek kiterjedten építenek a social engineering használatára, beleértve ebbe a deepfake használatát.

A deepfake mint eszköz egyre fontosabb megjelenését támasztja alá a social engineering támadások során az Európai Kiberbiztonsági Ügynökség (European Union Agency for Cybersecurity, ENISA) 2021-ben kiadott, aktuális kiberbiztonsági fenyegetéseket tartalmazó jelentése, az *ENISA Threat Landscape 2021* is. Ugyan ez a kiadvány még a háború előtt jelent meg, de figyelmeztetései a háborús időszakban is megállják a helyüket, hiszen összesen hét helyen említi a deepfake jelentette veszélyeket, és ezek a fenyegetések kivétel nélkül relevánssá váltak a gyakorlatban is. Az aktuális trendek közül kiemeli, hogy az államilag támogatott csoportok felhasználhatják ezt a technológiát bizonyos személyek megszemélyesítésére az információs műveletek során, mely műveletek közé egyébként a PSYOPS is tartozik. A részletes elemzésben a dezinformációk kapcsán említi meg az anyag azt, hogy a deepfake lehet ezek egyik kiemelt eszköze, mivel a technológia villámgyorsan fejlődik, a közösségi média pedig szélsebesen tudja terjeszteni, hatékony ellenintézkedések pedig nem igazán vannak. Emellett az online csalásoknál

is találkoztak vele a kutatók, konkrétan a Covid-19 tematikájú álwebshopoknál használták a bizalomébresztés céljából létrehozott deepfake-videókat, amivel a támadók célja minél több személyes és bankkártyaadat megszerzése volt (ENISA 2021).

A deepfake tehát valós kiberbiztonsági kockázatot jelent, így célszerű tételesen áttekinteni, hogy a négy fő fenyegetési kategória, azaz a kiberbűnözés, a kiberkémkedés, a hacktivizmus és kiberterrorizmus, valamint a kiberhadviselés során használták-e már ezt a technológiát, illetve a szakirodalomban lehetségesnek tartják-e későbbi felhasználását. A fejezet hipotézise szerint minden esetben, amikor a pszichológiai manipuláció hatékony lehet egy kibertámadás kivitelezésében, a deepfake felhasználható, és vagy van erre konkrét esettanulmány, vagy a szakirodalmi kutatás segítségével alátámasztható a valós kockázat.

2. A MESTERSÉGES INTELLIGENCIA JELENTETTE ÁLTALÁNOS KIHÍVÁSOK

Még mielőtt azonban elmerülnénk a deepfake mélységeiben, érdemes feltenni a kérdést, miért tekintjük valós veszélynek éppen ezt a technikát. A válasz pedig az, hogy nem kizárólag a deepfake, hanem általánosságban a mesterséges intelligencia felhasználásának robbanásszerű terjedése tölti el aggodalommal a kiberbiztonsági szakértőket. A klasszikus kibervédelmi paradigmák alapvetően arra építenek, hogy ismerjük a támadó technikáinak jelentős részét, amelyek aránylag lassan és kiszámíthatóan fejlődnek, tehát ezekre lehet olyan szabályalapú védelmet építeni, amelynek segítségével a támadónak jelentős költséget jelent a védelmi intézkedések kijátszása. Az igazán kritikus információs rendszerekben ezen elv alapján fejlesztették a védelmi megoldásokat az elmúlt évtizedekben. Az incidensek száma azonban exponenciálisan emelkedett, így a klasszikus, szabályalapú védelem kivitelezhetetlenné vált, de a kiberbiztonsági fenyegetéssel kapcsolatos rengeteg, rendelkezésre álló információ *(cyberthreat intelligence, CTI)* lehetővé tette, hogy a gépi tanulás útján a védelem egyre jobban kiismerje a támadási metódusokat. Ezzel párhuzamosan viszont a támadók is elkezdték kiaknázni a mesterséges intelligenciában rejlő lehetőségeket, és olyan támadási modelleket alkottak, amelyekkel szemben a klasszikus védelmi metódusok tehetetlenek. Így a 2020-as évekre egyre inkább azt tapasztaljuk, hogy mesterséges intelligencia által segített támadásokat próbálnak mesterséges intelligencia által kivédeni.

Ennek leglátványosabb megjelenése a képi információk feldolgozása és mesterséges intelligencia által történő újraalkotása. Mivel az interneten hihetetlen mennyiségű képi és videóanyag található, az MI-algoritmusok tanítása könnyedén megoldható, olyannyira, hogy az olyan nagy adataggregáló cégek, mint a Google jelentősen hozzájárulnak a mesterséges intelligencia tudományágához (Ekler–

Pásztor 2020). Az általuk kezelt adatmennyiség segítségével pedig sorra lépnek piacra azok a szolgáltatók, amelyek segítségével egy kibertámadás sikeresen előkészíthető. Így korábban csak a thispersondoesnotexist.com oldalon létrehozott, valódihoz nagyon hasonló arcképektől kellett tartani, amelyeket kiválóan lehet vegyíteni a fakepersongenerator.com oldalon előállított hamis személyazonosságokkal, jelenleg azonban néhány dollárért elérhetőek azok az arccserélő (FaceSwap) programok, amelyekkel gond nélkül le lehet forgatni egy videót úgy, hogy arra a generált arckép kerül. Sőt, a népszerű *American Idol* című tévéműsorban már olyan deepfake-videót is lehet látni, amelyben élőben illesztik az egyik műsorvezető arcát a videóképre, ezzel előrevetítve a kibertéri csalások új dimenzióját (Tangermann 2022).

Naik és szerzőtársai (2021) elemző cikkükben áttekintik, hogy milyen kihívásokat jelent a mesterséges intelligencia felhasználása a kibervédelemben. Egyik megállapításuk szerint a kiberbűncselekmények elkövetői, például a hackerek, alkalmazkodó támadások modellezésével és intelligens rosszindulatú programok létrehozásával mesterségesintelligencia-technikákat használhatnak a biztonsági intézkedések kikerülése során. Az ilyen programok információt gyűjthetnek arról, hogy mi akadályozta meg a támadásokat, majd ezek alapján megtanulják a későbbi támadások sikeres végrehajtását és az önterjesztést. A hackerek a mesterségesintelligencia-technológiákat olyan rosszindulatú kártékony kódok létrehozására is használhatják, amelyek képesek a megbízható rendszerelemek utánzására. Ennek kivédésére azt javasolják, hogy a védelemért felelős szervezetek az elkövetőknél magasabb szintű és optimális mesterségesintelligencia-módszereket használjanak a kiberbiztonságban. Például a mesterséges intelligenciával támogatott automatizált hálózat- és rendszerelemzés jó megoldás lehet a támadók által a rendszer védelmének áthatolására használt vezetési és irányítási *(command and control, C2)* taktikák elleni küzdelemben. Az ilyen automatizált adatkezelés biztosítja a rendszerek folyamatos felügyeletét a támadási kísérletek gyors azonosítása érdekében. További javaslatuk egy olyan mesterségesintelligencia-alapú kiberbiztonsági rendszer, amely a felhasználói interakciók előzményei alapján dolgozhat, és következtethet a várható viselkedésre, amelyet nehéz kihasználni (Naik et al. 2021). A gyakorlatban azonban a deepfake ellen csak elméleti vagy kísérleti megoldások léteznek, és ezeket sem használják tömegesen. Megállapíthatjuk tehát, hogy a deepfake az a mesterségesintelligencia-megoldás, amelyhez hozzáférni, a művelet során felhasználni könnyű, védekezni ellene azonban nagyon nehéz. Nem véletlen az egyre szélesebb körű alkalmazása még az olyan támadások során is, amelyek nem különösebben bonyolultak, és nem fontos célpontok ellen irányulnak.

3. DEEPFAKE A KIBERBŰNÖZÉSBEN

A deepfake felhasználásával a gyakorlatban leginkább a kiberbűnözés során lehet találkozni. Ez nem véletlen, hiszen elsősorban az online csalások során a már említett social engineering támadást lehet kiválóan automatizálni a deepfake-technológiával. Az elmúlt években egyre több olyan csalás és más kiberbűncselekmény derült ki, ahol ezt a megoldást használták. A teljesség igénye nélkül néhány nagyobb sajtóvisszhangot kiváltó eset:

- **CEO vagy Business E-mail Compromise (BEC) csalások**: Ebben a csalástípusban a támadó üzleti vezetőnek vagy üzleti partnernek adja ki magát, és a pénzügyi felelősnek ad utasításokat arra vonatkozóan, hogy valamilyen sürgős pénzügyi tranzakciót hajtson végre. Ez hagyományosan e-mail alapon történik, de vannak olyan esetek, amelyek során az utasítás szóban, telefonon keresztül érkezett (Mezei 2019: 32). A legelső ilyen esetről 2019-ben számoltak be a lapok, amikor egy brit energetikai cég képviselője gondolta azt, hogy német főnökével beszél, aki arra utasította, hogy egy jelentősebb összeget utaljon át az egyik állítólagos magyarországi beszállítónak. A támadó valójában egy hangszintetizátort használt (Stupp 2019).
- **Kriptovalutás csalások**: A covid-pandémia időszakában nagy népszerűségre tettek szert a kriptovaluták, amelyek kereskedése az egekbe szökött 2021-ben. Sokan kiváló befektetési lehetőséget láttak benne, és ez felkeltette a csalók figyelmét is. Az online fórumokon rendre olyan hirdetésekbe lehetett botlani, amelyek a különböző kriptovalutákat ajánlották a felesleges pénzzel rendelkezőknek. Egyes hirdetésekben látszólag világszerte ismert hírességek tűntek fel, természetesen hamis módon. A deepfake-technológia segítségével hozták létre ezeket a videókat, amelyekben például a brit milliárdos, Richard Branson „ajánlotta" a kriptotermékeket (Akhtar 2022). Az előre megkreált videókhoz képest előrelépés lehet a valós idejű deepfake használata. A legnagyobb kriptovaluta-platform, a Binance egyik vezetője, Patrick Hillmann arcát egy olyan családhoz használták, amelyben a bűnözők a Binance nevében tárgyaltak a Zoom platformon keresztül más szervezetek képviselőivel (Hurst 2022). Amennyiben ez valóban így történt, márpedig az elérhető technológia alapján ez nem elképzelhetetlen, 2021-ről 2022-re érdemben lépett előre a deepfake felhasználása az online csalásokban, és váltottak a bűnözői körök a valós idejű videómanipulációra.
- **Hírességek lejáratása**: A celebritások képmásait nemcsak a csalásokban használják fel, hanem sokszor egyszerűen szórakozásból is készülnek olyan lejáratóvideók, amelyekben látszólag egy jól ismert arc csinál vállalhatatlan dolgot. Mivel az online térben számos olyan videó kering, amely alapul szolgálhat a gépi tanulást használó deepfake-algoritmusoknak, a közösségi média tele van olyan kamufiókokkal, amelyek egy híresség alteregójának

építettek. Ilyen például a Tom Cruise amerikai színészt megidéző @deeptomcruise TikTok-fiók, 2022 őszén 3,6 millió feliratkozóval (Cover 2022). Az ismert emberek imázsa sokszor dollármilliókat ér, így nem elképzelhetetlen egy olyan zsarolássorozat, amelyben a bűnözők pénzt kérnek azért, hogy ne indítsanak lejáró kampányt egy-egy jól forgó sztárral szemben. Könnyen lehetne ugyanis olyan helyzetet teremteni, amelyben az áldozat látszólag olyan dolgot tesz, ami a #meetoo utáni világban a karrier azonnali kettétörésével járna.

- **Hamis pornográf videók, bosszúpornó**: A deepfake első széles körű felhasználása a híresség arcának ráillesztése volt nyíltan pornográf videókra. A Deeptrace nevű cég 2019-es kutatása alapján az akkor fellelhető deepfake-videók 96%-a volt pornográf, és ezek 99%-ában egy női híresség arcát használták fel (Sample 2020). Az erre vonatkozó igény nem változott, egy egyszerű Google-kereséssel a mai napig számos olyan oldalt lehet találni, ahol celebritások deepfake-pornóit lehet megnézni. Mivel a gépi tanuláshoz szükséges adatok rendelkezésre állnak egy híresség esetén, felmerül a kérdés, hogy egy átlagember esetében vajon megvalósítható-e ugyanez. 2021 elején került elő ugyanis az az eset, ami szerint egy pennsylvaniai édesanya állítólagosan deepfake-videót készített a lánya cheerleader-csapatának egyik tagjáról, hogy ezzel járassa le őt. A pornográf tartalomról utólag kiderült, hogy az valódi, nem szerkesztett, és a helyi rendőrség hibájából terjedt el az állítólagos kreált videó története, de a technológia fejlődése és az arccserét lehetővé tevő „faceswap" alkalmazások elérhetősége ma már komolyan fenyegeti az átlagembert is azzal, hogy hamis felnőttvideót hoznak róla létre az ellenlábasai (Ho Tran 2021).

A deepfake elterjedése nemcsak az elkövetésben segíti a bűnözőket, hanem a védekezésben is. Gondoljunk csak bele abba, hogy ha az áldozat egy hamis videóról elhisz bármit, akkor egy valódi videóról miért ne állíthatná az elkövető, hogy az tulajdonképpen manipulált, és nem is ő látható a felvételen. Ez történt 2021 novemberében, amikor egy ismert magyar youtubert gyanúsítottak meg azzal, hogy egy online csetelés során rávett egy 11 éves kislányt arra, hogy levetkőzzön a kamerák előtt (Barnóczki 2021). A megvádolt fiatalember a nála tartott házkutatás után közzétett videóján hosszasan magyarázta, hogy szerinte a Discord csatornájához szintén hozzáférő barátai hamisították meg a hangját és képét, ezt bizonyítandó pedig (igencsak kevéssé meggyőző) videófelvételt is bevágott. A fiatal minden bizonnyal a deepfake-et próbálta a saját védelmére felhozni, de a konkrét esetben valószínűsíthetően sok egyéb bizonyíték is ellene szólt, ami megalapozta a meggyanúsítását. A rendőrségnek azonban ezek után komolyan számolnia kell azzal, hogy a gyanúsított megkérdőjelezi egyes bizonyítékok hitelességét, ami új típusú rendészeti kihívást jelenthet.

Összességében tehát a deepfake megoldások széles körben rendelkezésre állnak a bűnelkövetők számára, és ez az eszköz 2021-re már beépült a kiberbűnözés eszköztárába. A technológia olcsón és könnyen elérhető, minden bizonnyal továbbra is széles körben fogják használni, miközben a potenciális áldozatok jelentős része nem is hallott erről a fenyegetésről. A bűnmegelőzéssel foglalkozó szakembereknek tehát komolyan kell foglalkozniuk ezzel a jelenséggel. Csakúgy, mint a szabályozóknak, hiszen ahogy Sorbán Kinga megállapítja, a „hozzájárulás nélkül közzétett pornográf felvételek és a hamisított felvételek által okozott probléma kétséget kizáróan valós. Az előfordulás számosságára és az áldozatoknak okozott kárra tekintettel e cselekmények büntetőjogi fenyegetettsége egyáltalán nem tűnik túlzó lépésnek" (Sorbán 2020: 101). Véleménye szerint az egyik legfontosabb teendő a deepfake-videóknak helyt adó platformok egységes szabályozása lenne, ami valóban jó, rendszerszintű válasznak tűnik a deepfake fenyegetésekkel szemben.

4. DEEPFAKE A KIBERKÉMKEDÉSBEN

A kémkedés szinte egyidős az emberiséggel, az információ pedig közismerten maga a hatalom, és mivel az információk ma már szinte kizárólag elektronikus formában keletkeznek, tárolódnak, majd kerülnek továbbításra és feldolgozásra, nem meglepő, hogy a világ titkosszolgálatai aktívan élnek a kibertéri hírszerzés lehetőségével. Ahogy a kiberbűnözés esetében, úgy a kiberkémkedésnél is a social engineering az egyik legfontosabb támadói módszer, így a deepfake a kiberhírszerzésnél is megjelenik. Mivel állami tevékenységről van szó, tipikusan célzottan felhasználva egy bizonyos személlyel szemben, a nyilvánosság elé nagyon kevés eset kerül ebből a körből.

A deepfake titkosszolgálati felhasználását a titkossága miatt ezért nehéz rekonstruálni, de az információmorzsák alapján ugyanazt a fejlődési ívet követi, mint amit a kiberbűnözésnél láttunk. Először a mesterséges intelligencia által generált arcok jelentek meg. Egy 2019-ben nyilvánosságra került akcióban egy magát Katie Jonesnak nevező személy kezdett el a LinkedIn szolgáltatáson keresztül kapcsolatokat kialakítani fontos *think tank*ek munkatársaival. Jones profilján meggyőző korábbi munkahelyek, kapcsolatai között pedig mérvadó szakemberek voltak, akik valószínűsíthetően automatikusan jelölték őt vissza. Ez elég hitelt adott személyének ahhoz, hogy egyre többen igazolták a vele való ismeretséget. Az álprofil ezután kezdett el direkt kapcsolatba lépni az új ismerősökkel, ami először Keir Gilesnek, a londoni Chatham House Oroszország-szakértőjének tűnt fel. Mivel Jones adatlapja szerint a washingtoni Center for Strategic and International Studies dolgozója, Giles pedig szinte mindenkit ismer onnan, gyanússá vált neki, hogy vajon miért nem találkozott a hölggyel korábban. Kis utánajárás után kiderült, hogy ez az információ hamis, csakúgy, mint a profilkép, amelyet a mestersé-

ges intelligencia hozott létre. Az álprofil ezután el is tűnt. A megcélzottak köre arra mutat, hogy ebben az esetben egy orosz műveletet sikerült kiszúrni, de a közösségi hálózatokon történő ilyen kapcsolatépítésben egyébként a kínai hírszerzők járnak az élen (Satter 2019).

Az állami hírszerző szervezetek mellett természetesen a magáncégeknek is fontos versenytársaik titkainak kifürkészése. Túllépve a mesterséges intelligencia által generált fényképeken, 2022 nyarán az FBI Internet Crime Complaint Center arra figyelmeztette a vállalatokat, hogy megnövekedett aktivitást érzékelnek az álláshirdetésekre jelentkező deepfake-személyiségek számában. A támadás konkrét lefolyása úgy néz ki, hogy egy cég meghirdet egy jellemzően informatikai pozíciót. Erre jelentkezik valaki egy hamis személyazonossággal. Az állásinterjút távolról, valamilyen internetes videókonferencia megoldáson keresztül bonyolítják le, ahol a valódi személy arca helyett egy deepfake jelenik meg, kihasználva a valós idejű leképezésben rejlő lehetőségeket. A támadó célja olyan informatikai munkakörben állást szerezni, amely lehetőséget biztosít a távoli munkavégzésen keresztül a szervezet informatikai rendszereihez való hozzáférésre. Ha pedig már bent van a hálózaton, semmi sem akadályozza meg abban, hogy a célszervezet titkaihoz hozzá tudjon férni. A kommentárok alapján ebben az esetben észak-koreai támadókat sejtettek a visszaélés mögött, de nem egyértelmű, hogy állami vagy nem állami hátterű támadásról van-e szó (FBI 2022).

A deepfake jelentette nemzetbiztonsági veszélyt támasztja alá Észtország Külföldi Hírszerző Szervezetének 2021-es éves jelentése, amelyben egy teljes oldalt szentelnek a Katie Jones-ügynek, kiemelve, hogy ez egy példa az orosz befolyásolási manőverekre (Estonia 2021). Ellentevékenységként a tudatosításra hivatkoznak, hiszen minél többen ismerik az ilyen hírszerző technikákat, annál nagyobb a valószínűsége annak, hogy egy célszemély azt észreveszi, és jelenti a belső elhárító szervezeteknek. Ez azonban feltételezi a lehetséges célszemélyek éberségét és legalább az alapvető technológiai támadások megértésének igényét. Sajnálatos módon azonban a legérzékenyebb területeken dolgozók jellemzően elzárkóznak ettől az általuk túlzottan is műszakinak gondolt területtől, és arra hivatkozva, hogy sem koruk, sem végzettségük nem megfelelő a kibertámadások értelmezéséhez, inkább nem is foglalkoznak a kérdéssel. Műszaki védelmet viszont szinte lehetetlen ezekre az esetekre építeni, így nincsen más megoldás, csak tudatosan beépíteni az oktatási anyagokba a deepfake-technika felismerésének trükkjeit is.

5. DEEPFAKE A HACKTIVIZMUSBAN ÉS A KIBERTERRORIZMUSBAN

Hacktivistának nevezzük azokat, akik egy politikai ideológia mentén szervezett kibertéri akcióban vesznek részt, amelynek hatása van a fizikai világra is. A hacktivista akciók eszköztára közel sem annyira összetett, mint amit egy kiváló műveleti tervező képességekkel rendelkező állami szereplő kivitelezni tudna. A hacktivista csoportok ereje ráadásul a láthatóságban van, így nem érdekeltek abban, hogy a művelet rejtve legyen, illetve tipikusan csoportosan követik el a cselekményt, sokszor egymást nem is ismerve, akár nagyon eltérő földrajzi helyekről. Ráadásul ezek a csoportok jellemzően bizonyos ügyek mentén szerveződnek, ad hoc módon, ami szintén csökkenti a szerveződés lehetőségeit. A konspiráció tehát nem feltétlenül cél, szemben a korábban tapasztalt esetekkel. Marco Romagna összefoglalója alapján így egy hacktivista cselekmény jellemzően az elosztott túlterheléses támadásokra (DDoS), a weboldalak átírására *(defacement)* és az adatlopásra korlátozódik. Időnként előfordulhat kártékony kódok alkalmazása, de ennek meglehetősen negatív visszhangja van a közösségben. Ezek között nem szerepel a social engineering, és nehezen párosítható hozzá a deepfake mint eszköz (Romagna 2019).

Az orosz–ukrán háború azonban rávilágított a hacktivisták egy speciális csoportjára, a patrióta vagy más néven hazafias hackerek közösségére, akik a katonai kiberműveletek szempontjából nagyon fontos csoportot alkotnak a hacktivistákon belül. Athina Karatzogianni hacktivizmusról szóló könyvében (2015) úgy írja körbe a patrióta hackereket, mint akik a nemzetük tisztaságáért küzdenek az online média ügyes felhasználásával. Paradox módon tehát a nacionalizmus mint politikai ideológia jelenik meg a klasszikus hacktivista akciók mögött, kihasználva az internetet mint globális médiát (Karatzogianni 2015: 22). Az orosz–ukrán háború folyamán ezek a csoportok mindkét ország mögött felsorakoztak, sőt, számos harmadik országban működő hacktivista csoport is beszállt a kibertéri küzdelmekbe, az egyik vagy a másik felet támogatva. Eszköztáruk pedig kiteljesedett, sokszor kimondottan destruktív, illetve a kiberkémkedés kategóriájába hajló cselekményeket hajtanak végre. Mivel a háború során tevékenységük nem egyértelműen elválasztható az államok által végrehajtott kibertéri műveletektől, így feltételezhető, hogy tevékenységüket a hadműveleti tervezők kontrollálják vagy akár irányítják is, tehát a deepfake hackereknek tulajdonított felhasználását a kiberhadviselés keretein belül kell vizsgálni.

A „békeidőben" végrehajtott hacktivista cselekmények esetében az irodalomkutatás során nem sikerült fellelni a deepfake-et felhasználó esetet. Logikusan gondolkodva azonban lenne ennek értelme, hiszen például az Anonymous-csoport jellegzetes, maszkos, eltorzított hangú figyelemfelhívó videói elkészíthetőek lennének ezzel a technikával. Emellett a nyilvánossággal való kapcsolattartás so-

rán is könnyebb lenne a teljes és biztos anonimitás fenntartása egy kreált perszónán keresztül. A célpont megszemélyesítése a mesterséges intelligencia által azonban már kétélű fegyver lehet, hiszen a hacktivisták hitelességét diszkreditálná, ha valótlanságot terjesztenének.

A hacktivizmus mellett szót kell még ejteni a kiberterrorizmusról is, hasonlóságuk okán. Cohen (2010) szerint a kiberterrorizmus a számítógép-hálózatok felhasználása az emberi élet kioltása vagy a nemzeti kritikus infrastruktúra szabotálása céljából, amelyet olyan módon követnek el, hogy az emberi életet veszélyeztethet. A kiberterrorizmus magában foglalja a számítógépek és/vagy a kapcsolódó technológia felhasználását azzal a szándékkal, hogy kárt vagy sérülést okozzanak a polgári lakosság kényszerítése és a célkormány politikájának, illetve magatartásának egyéb módon történő befolyásolása érdekében. Továbbá a kiberterrorizmus – amelyet meg kell különböztetni a hacktivizmustól és a kiberháborútól – magában foglalja a kritikus infrastruktúrák elleni támadást. A kiberterrorizmus és a terrorizmus között tágabb értelemben véve is vannak kapcsolódási pontok és eltérések, amelyek mindkét esetben elkerülhetetlenül befolyásolják a terrorizmus elleni válaszlépéseket. A kiberterrorizmus tehát felülről történő megközelítésben a „hagyományos" terrorizmus (illetve gerilla-hadikultúra) céljainak infokommunikációs eszközökkel való végrehajtása vagy alulról tekintve az informatikai bűncselekmények tömeges, szervezett végrehajtása. A kiberbiztonsági szakma nagyjából egységes álláspontja alapján kiberterrorista cselekmény nem vált ismertté, bár az egyes terrorcsoportok aktívan használják a kiberteret és a digitális technológiákat. Ennek megfelelően konkrét deepfake-kel kapcsolatos eseményt sem ismerünk. A terrorszervezetek propagandájában és a tagok beszervezésében, instruálásában azonban ez a technológia is megjelenhet, ezt azonban nem kiberbiztonsági fenyegetésként értelmezzük.

6. DEEPFAKE A KIBERHADVISELÉSBEN: INFORMÁCIÓS MŰVELETEK A HIBRID HADVISELÉSTŐL A KONKRÉT HÁBORÚIG

A mesterséges intelligencia, ezen belül pedig a deepfake használata régóta foglalkoztatja a katonai tervezőket, mivel a technológiában hihetetlen potenciál van, amely kiválóan tudja segíteni az olyan klasszikus katonai tevékenységeket, mint az információs műveleteket, ezen belül is a lélektani műveleteket. Bányász Péter és szerzőtársai (2019) rávilágítanak arra, hogy az információs műveleteket támadó és védelmi célból egyaránt lehet használni, céljuk pedig az, hogy hatást váltsanak ki gazdasági, politikai és katonai alrendszerekben, vagyis hogy el lehessen érni az információs fölényt, amely hozzájárul a győzelemhez. A Magyar Honvédség által, 2014-ben kiadott *Információs műveletek doktrína* alapján: „A Lélektani Műveletek (PSYOPS) elsődleges célja, hogy befolyásolja egy kiválasztott célcsoport viselke-

dését, magatartásformáit és véleményét az elöljáró által elfogadott PSYOPS célokkal összhangban, valamint hogy kiváltsa vagy megerősítse a célcsoport kívánt viselkedését az elöljáró távlati céljainak érdekében". A közösségi média mint terjesztő közeg és az erről tájékozódó tömegek, beleértve a döntéshozókat mint célcsoportot kiválóan alkalmas arra, hogy a korábban megismertek alapján a deepfake-technika segítségével manipulálni lehessen a „szíveket és az agyakat", azaz az információs tér kognitív dimenzióját.

Ezzel a veszéllyel évek óta tisztában vannak a nemzetvédelemért felelős személyek. Mivel elsősorban Oroszország, de más országok is évek óta használják a kiberteret a „se nem béke, se nem háború" idején szokásos hibrid műveletek végrehajtására, komoly tapasztalatot lehetett szerezni azzal kapcsolatban, hogyan próbálják egyes országok békeidőben befolyásolni a közvéleményt a közösségi médián keresztül. A deepfake ebben a körben elsősorban a választások manipulálásával kapcsolatban került szóba. Az Egyesült Államokban komolyan felvetődött annak a kockázata, hogy külső szereplők deepfake segítségével fogják befolyásolni a 2020-as elnökválasztást. Ezt szerencsére sikerült elkerülni, mindössze két kampányfilmet készítettek ezen a módon, beazonosíthatóan belföldi szereplők. A támadások kivédésében azonban nagy szerepe lehetett azoknak a platformoknak, amelyek az amerikai belbiztonsági szervezetekkel együttműködve aktívan vadásztak a deepfake-alapú információs kampányokra (Simonite 2020).

Háborús időszakban azonban egészen az orosz–ukrán háborúig nem volt példa a deepfake használatára. Oroszország információs műveleteit tanulmányozva egyértelmű volt, hogy előbb-utóbb ezt a fegyvert is bevetik. Erre nem kellett sokat várni, 2022. március 16-án, három héttel a háború megindulása után meg is jelent az a videó, amelyben Volodimir Zelenszkij elnök bejelenti Ukrajna fegyverletételét. Ezt megerősítendő, az Ukrayina 24 nevű tévécsatorna weboldalán is megjelent erről egy hír. A videó természetesen hamis volt, a weboldalt pedig feltörték. A hír gyorsan kezdett el terjedni a közösségi médiában, olyannyira, hogy Zelenszkij elnöknek személyesen kellett ezt cáfolnia, illetve az ukrán stratégiai kommunikáció is erőteljesen hangoztatta a videó valótlanságát. Ezzel sikerült elejét venni a komolyabb hatásnak (Simonite 2022). Válaszul viszont nagyon hamar megjelent egy hasonló tartalmú videó Vlagyimir Putyin elnökről, aki pedig az orosz katonák fegyverletételét jelenti be a kreált felvételen. Természetesen ennek a cáfolata is hamar megjelent, de ez a két eset is rávilágított arra, hogy mennyire fontos a hadban álló felek számára, hogy legyenek forgatókönyveik a deepfake-videókra (Baig 2022).

A felkészültség tehát kulcsfontosságú. A választási kampányok védelmére a Carnegie Endowment for International Peace kiadott egy útmutatót, amely a legfontosabb lépéseket fogalmazza meg a kampányban részt vevő pártok és jelöltek számára:

- Adjon ki nyilatkozatot arról, hogy a jelölt tudatosan nem fog hamis vagy manipulált híreket terjeszteni ellenfeleiről, és felszólítja a kampány támogatóit, hogy ugyanezt a kötelezettségvállalást tartsák be.
- Ismerje meg a közösségimédia-platformok szolgáltatási feltételeit és közösségi irányelveit ebben a kérdésben, valamint legyen tisztában a nem megfelelő tartalmak bejelentésére szolgáló eljárásokkal.
- Jelöljön ki egy csapatot, amely készen áll az incidens kezelésére.
- Kérjen tájékoztatást a legfontosabb trendekről és fenyegetésekről szakértőktől.
- Végezzen belső „red teaming" gyakorlatot, felkészülve arra, hogy egy hamisítvány milyen módon veheti célba a jelöltet vagy a kampányt.
- Alakítson ki kapcsolatokat azon fontosabb szervezetek képviselőivel, amelyek hasznosak lehetnek egy incidens során.
- Legyenek olyan eljárások, amelyek segítségével gyorsan hozzá lehet férni a hamisítvány alapját képező eredeti videó- és/vagy hangfelvételekhez.
- Készítsen elő olyan vészhelyzeti webes tartalmakat vagy sablonokat, amelyeket gyorsan fel lehet használni a hamis állítások ellen (Charlet–Citron 2019).

Ez működőképes lehet egy politikai kampány során, de háborús szituációban ennél strukturáltabb védelemre van szükség. A 2017-ben Helsinkiben közös NATO–EU kezdeményezésre alapított Hibrid Fenyegetések Elleni Európai Kiválósági Központ (European Centre of Excellence for Countering Hybrid Threats) által kiadott, az orosz–ukrán háborúban alkalmazott deepfake akciókat értékelő elemzés például a következő megállapítással zárul: „A 2022-es ukrajnai háború azt mutatja, hogy a deepfake használata egy feltörekvő trend, és még ha a sugárzott videók meglehetősen kidolgozatlannak tűntek is, nem szabad elfelejteni, hogy még csak az MI-alapú befolyásolás korszakának előestéjén járunk. A NATO-nak, az EU-nak és az európai országoknak ezért fontolóra kell venniük az információs területet használó befolyásolási és befolyáselhárítási műveletek doktrínájának és folyamatának jelentős frissítését. Az ukrajnai deepfake-videók felhasználásával felmerült annak szükségessége, hogy a GAN-technológiák segítségével belső képességeket kell kifejleszteni a deepfake felderítésére és elhárítására. A közeljövőben tehát az MI-alapú technológiák alkalmazása hamis beszéd és videók létrehozására új normává válhat a hagyományos és nem hagyományos hadviselésben. Az információs hadviselés 3.0 elleni küzdelemhez szükséges ismeretek megszerzéséhez elengedhetetlen az erre szakosodott személyzet képzése. Mivel a befolyásolással kapcsolatos képzések nagy része történelmileg a pszichológia és a hagyományos média használata körül forgott, a mesterséges intelligencia alapú technológiák növekvő jelentősége miatt most egy új réteggel kell ezeket bővíteni. Mivel ezek a technológiák nem csak a hadsereghez kapcsolódnak, és természetüknél fogva nem tekinthetők »katonai felszerelésnek«, valószínűleg nem lesz jogi lehetőség a használatuk korlátozására" (Mazzucchi 2022).

7. KONKLÚZIÓ

A szakirodalmi kutatás megerősítette, hogy a pszichológiai manipuláció egyik fontos és egyre fontosabb eszköze a deepfake, amelyet használnak célzottan, egyes személyek ellen, nagyobb körben, csalások során vagy akár teljes társadalmat célzóan, az információs műveletek részeként. Az első konkrét esettanulmányok médiában történő említései 2019-ben jelentek meg, ekkor vált a technológia széles körben elérhetővé, ekkoriban viszont még elsősorban az MI által generált arcok felhasználása volt a középpontban. A covid-pandémia alatt jelentek meg azok a szoftverek, amelyekkel már korábban rögzített videókat is manipulálni lehetett, 2022-ben pedig már akár a valós idejű arccsere is megvalósítható volt.

Ezt a nagyon gyors technológiai fejlődést a védelmi oldal nem tudta követni, a széles felhasználói réteg egyrészt nincsen tudatában a deepfake jelentette kockázatoknak, másrészt nincsenek széles körben használható védelmi megoldások. A kibervédelem oldaláról három lehetséges ellenintézkedés mutatkozik. Ezek közül az első a témával kapcsolatos általános tudatosság növelése. Amennyiben a lehetséges áldozatok tudatában vannak annak, hogy a deepfake valós veszélyt jelent, lehetőség mutatkozik a hatékonyabb védelemre. Második megoldásként a digitális platformszolgáltatók szabályozáson keresztüli kényszerítése jelenthet megoldást, aminek folyományaként olyan technológiai megoldások születhetnek, amelyek hatékonyan zárják el a deepfake-től a tömegeket. Harmadikként pedig az államok stratégiai kommunikációjának kell tudnia kezelni azt a helyzetet, amikor hibrid műveletként vagy katonai konfliktus keretében idegen államok használják ezt a megoldást. Ezek mindegyikének meg kell valósulnia az állami és európai kibervédelem keretében, így bár a deepfake még néhány évig jelen lesz érdemi kockázatként, az összehangolt védekezés valószínűsíthetően csökkenteni fogja az ezt használó támadások hatását.

SZAKIRODALOM

Bányász Péter – Dobos László – Palla Gergely – Pollner Péter 2019: Lélektani műveletek a közösségi médiában. In: Auer Ádám – Joó Tamás (szerk.): *Hálózatok a közszolgálatban*. Budapest: Dialóg Campus Kiadó. 111–134.

Charlet, Katherine – Citron, Danielle 2019: Campaigns Must Prepare for Deepfakes: This Is What Their Plan Should Look Like. *Carnegie Endowment for International Peace*, szeptember 5. https://carnegieendowment.org/2019/09/05/campaigns-must-prepare-for-deepfakes-this-is-what-their-plan-should-look-like-pub-79792 [2022. 09. 23.]

Cohen, Aviv 2010: Cyberterrorism: Are we legally ready? *Journal of International Business and Law*, 9/1: 1–40.

Deák Veronika 2019: Social engineering alapú információszerzés a kibertérben megvalósuló lélektani műveletek során. *Hadtudományi Szemle*, 12/3: 95–111. doi: 10.32563/hsz.2019.3.6

Ekler Péter – Pásztor Dániel 2020: Alkalmazott mesterséges intelligencia felhasználási területei és biztonsági kérdései. Mesterséges intelligencia a gyakorlatban. *Scientia et Securitas*, 1/1: 35–42. doi: 10.1556/112.2020.00006

Karatzogianni, Athina 2015: *Firebrand Waves of Digital Activism 1994–2014: The Rise and Spread of Hacktivism and Cyberconflict*. Basingstoke: Palgrave Macmillan.

Mazzucchi, Nicolas 2022: AI-based technologies in hybrid conflict: The future of influence operations. *Hybrid CoE*, június. https://www.hybridcoe.fi/wp-content/uploads/2022/06/20220623-Hybrid-CoE-Paper-14-AI-based-technologies-WEB.pdf [2022. 09. 23.]

Mezei Kitti 2019: A kiberbűnözés szabályozási kihívásai a büntetőjogban. *Ügyészek Lapja*, 26/4–5: 21–33.

Mitnick, Kevin D. – Simon, William L. 2003: *A legendás hacker. A megtévesztés művészete*. Budapest: Perfact-Pro.

Naik, Binny –Mehta, Ashir – Yagnik, Hiteshri et al. 2022: The impacts of artificial intelligence techniques in augmentation of cybersecurity: a comprehensive review. *Complex & Intelligent Systems* 8: 1763–1780. doi: 10.1007/s40747-021-00494-8

Oroszi Eszter Diána 2019: Social engineering technikák. In: Deák Veronika (szerk.): *Célzott kibertámadások. Éves továbbképzés az elektronikus információs rendszer biztonságával összefüggő feladatok ellátásában részt vevő személy számára 2018*. Budapest: Nemzeti Közszolgálati Egyetem. 77–118.

Romagna, Marco 2019: Hacktivism: Conceptualization, Techniques, and Historical View. In: Holt, T. – Bossler, A. M. (szerk.): *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. Palgrave Macmillan, Cham. doi: 10.1007/978-3-319-90307-1_34-1

Sorbán Kinga 2020: A bosszúpornó és deepfake-pornográfia büntetőjogi fenyegetettségének szükségességéről. *Belügyi Szemle*, 68/10: 81–104. doi: 10.38146/BSZ.2020.10.4

FORRÁSOK

Akhtar, Tanzeel 2022: Celebrity-Endorsed Crypto Scams Soaring in UK, Santander Says. *Bloomberg*, június 28. https://www.bloomberg.com/news/articles/2022-06-28/celebrity-endorsed-crypto-scams-soaring-in-uk-santander-says [2022. 09. 23.]

Baig, Rachel 2022: Fact check: The deepfakes in the disinformation war between Russia and Ukraine. *Deutsche Welle,* március 18. https://www.dw.com/en/fact-check-the-deepfakes-in-the-disinformation-war-between-russia-and-ukraine/a-61166433 [2022. 09. 23.]

Barnóczki Brigitta 2021: Gyermekpornográfia miatt nyomoz egy magyar youtuber után a rendőrség. *Telex,* november 19. https://telex.hu/belfold/2021/11/19/gyermekpornografia-miatt-nyomoz-egy-magyar-youtuber-ellen-a-rendorseg [2022. 09. 23.]

Cover, Rob 2022: Celebrity deepfakes are all over TikTok. Here's why they're becoming common – and how you can spot them. *The Conversation,* július 18. https://theconversation.com/celebrity-deepfakes-are-all-over-tiktok-heres-why-theyre-becoming-common-and-how-you-can-spot-them-187079 [2022. 09. 23.]

Estonian Foreign Intelligence Service 2021: *International Security and Estonia 2021*. Tallinn.

European Union Agency For Cybersecurity 2021: *ENISA Threat Landscape 2021*. Athens: ENISA.

Federal Bureau of Investigation 2022: Deepfakes and Stolen PII Utilized to Apply for Remote Work Positions. *FBI*, június 28. https://www.ic3.gov/Media/Y2022/PSA220628 [2022. 09. 23.]

Ho Tran, Tony 2021: Remember that deepfake cheerleader story? Turns out it was probably nonsense. *Futurism*, szeptember 28. https://futurism.com/the-byte/deepfake-cheerleader-video-nonsense [2022. 09. 23.]

Hurst, Luke 2022: Binance executive says scammers created deepfake 'hologram' of him to trick crypto developers. *Euronews*, augusztus 24. https://www.euronews.com/next/2022/08/24/binance-executive-says-scammers-created-deepfake-hologram-of-him-to-trick-crypto-developer [2022. 09. 23.]

Sample, Ian 2020: What are deepfakes – and how can you spot them? *The Guardian*, január 13. https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them [2022. 09. 23.]

Satter, Raphael 2019: Experts: Spy used AI-generated face to connect with targets. *AP News*, június 13. https://apnews.com/article/ap-top-news-artificial-intelligence-social-platforms-think-tanks-politics-bc2f19097a4c4fffaa00de6770b8a60d [2022. 09. 23.]

Simonite, Tom 2022: A Zelensky Deepfake Was Quickly Defeated. The Next One Might Not Be. *Wired*, március 17. https://www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/ [2022. 09. 23.]

Simonite, Tom 2020: What Happened to the Deepfake Threat to the Election? *Wired*, november 16. https://www.wired.com/story/what-happened-deepfake-threat-election/ [2022. 09. 23.]

Stupp, Catherine 2019: Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal*, augusztus 30. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 [2022. 09. 23.]

Tangermann, Victor 2022: Simon Cowell shocked by deepfake copy of himself. *Futurism*, június 9. https://futurism.com/the-byte/simon-cowell-deepfake [2022. 09. 23.]

JOG

LENDVAI GERGELY FERENC

Deepfake a szólásszabadság tükrében – reflexiók a jog perspektívájából

A fejezet a deepfake jogi szabályozását vázolja fel, különös tekintettel a szólásszabadságra. A fogalmi alapok után ismertetjük a jogellenes deepfake-tartalmak kontextusában releváns amerikai, európai és további szabályozásokat, illetve megoldási javaslatokat. A fejezet különös figyelmet szentel a véleménynyilvánítás és a szólásszabadság gyakorlásának kapcsolatára a deepfake-technológiával, valamint e kapcsolat jogi megítélésére, és nemzeti és nemzetközi szabályozások mellett több „kulcsesetként" számontartott példát is bemutat.

Kulcsszavak: szólásszabadság, mesterséges intelligencia szabályozása, esetjog

1. FOGALMI ALAPOK

A deepfake fogalmának meghatározása igen problémás (Ambrus 2021), különösen, ha a jog terminológiáját használjuk. A definíciót az érthetőség kedvéért érdemes egységekre bontva külön elemezni. Jellemzően a deepfake:

- egy olyan gépi tanulást, mesterséges intelligenciát, illetve mesterséges tanulást alkalmazó technológiát (Pantserev 2020: 40; G. Karácsony 2022: 301) vagy algoritmusokkal dolgozó technológiai megoldást jelent, amellyel
- a felhasználó célzatosan, szándékosan (Fernandez 2022; Paris–Donovan 2019; Ferraro 2019: 1–3) meg tud hamisítani (Mráz 2021: 249), valamiféle módon meg tud másítani vagy át tud alakítani (Lyu 2020: 1) – legtöbb esetben, de nem kizárólagosan emberekről készült (Cover 2022: 609) – kép- és hangfelvétel jellegű tartalmakat (Van der Sloot – Wagensveld 2020: 1; Veszelszki 2022: 33);
- úgy, hogy ezzel egy olyan hamisított, új tartalmat hoz létre, amely meggyőző módon hiteti el, hogy az új tartalom és azoknak szereplő(i) valóságosak, illetve eredetiek (Gibson 2020: 260; Ambrus 2021).

A szövegben a kötet címének megfelelően következetesen a „deepfake" szót használom, mivel a magyar és európai jogi fordítások (például: „mélykamu"; EPRC 2020: 88) nem adják át elég szemléletesen az angol szóösszetétel szemantikai tartalmát.

A deepfake-technológia „egy személyről rendelkezésre álló képek vagy hangfelvételek alapján olyan manipulált felvételt készít, amely a valós személyt egy olyan fiktív helyzetben ábrázolja, amilyen helyzetben a konkrét személy nem szerepelt, illetve olyan kijelentést tulajdonít neki, amilyen kijelentést ő nem tett meg" (G. Karácsony 2022: 306). Ugyanezt a technológiát használják gyakran a deepfake-videók „deepfake-jellegének" felderítésére is (vö. a ProofMode vagy a TruePic felismerő programok; Tariq et al. 2021: 3626–3632; Parentsev 2020: 40). A mesterséges intelligencia kiemelkedő szerepet játszik nemcsak a tartalmak alakításában, hanem a tartalmak valósághűségének fejlesztésében is. Ehhez különösen nagy mennyiségű adatra van szüksége, éppen ezért az, hogy a közösségi médián és egyéb platformokon a felhasználók önszántukból, maguk töltenek fel kép- és hangfelvételeket, ezzel folyamatos adatinputot szolgáltatva a deepfake-technológiát használó szoftvereknek, hatványozottan és eddig soha nem látott gyorsasággal fejleszti és teszi egyre valódibbá a deepfake-tartalmakat (G. Karácsony 2022: 307; Mráz 2021: 251–252).

Jelenleg három nagy arcalakítási deepfake technika ismert: 1. A „báb" technológiával egy eredeti személy arcától válláig terjedő testfelület és annak mozgása másolható a célszemélyre. Így a testi megnyilvánulásokon túl a mozgásokat is „bábmesterként" átmásolja a technológia. 2. Az arcmásolás *(face-swapping)* talán a deepfake legismertebb manifesztációja; ebben az esetben az arckifejezések, a mimika másolása történik (vö. Cover 2022: 609; Mezei–Szentgáli-Tóth 2022: 248 definíciói csak erre terjednek ki). 3. Az ajakszinkronizálás *(lip-sync)* segítségével az ajkak mozgásának áttételével lehet a célszemély közlését manipulálni (Lyu 2020: 301: 1–2).

Talán kevésbé egyértelmű a technológia által lefedett tartalom köre. Összességében a deepfake-tartalmak legtöbbször emberekről, különösen emberek arcáról és/vagy hangjukról készülnek (G. Karácsony 2022: 301; Lyu 2020: 1). Az arc digitális lemásolása és transzformációja két okból is jelentős a szólásszabadság és a jog kontextusában. Elsősorban az arc az egyén, a személy egyik legfontosabb testi jellemzője, egy azonosítási célra alkalmas evolúciós jellegzetesség (G. Karácsony 2022: 301). Másodsorban az arcot ábrázoló kép- és hangfelvételek mint médiumok különösen jelentősek a közbizalom, a közbeszéd és a tartalmak hitelessége kapcsán (Mráz 2021: 257; Mezei–Szentgáli-Tóth 2022: 248). Utóbbi a szólásszabadság és a demokratikus nyilvánosság tekintetében is kulcsfontosságú; a közlési hitelesség, illetve a közlés valódiságába vetett bizalom a társadalmi-politikai viszonyok kialakításának (lásd politikai közbeszéd), a demokratikus jogállamiságnak, a szabad véleménynyilvánítás és szólásszabadság érvényesülésének alapja (G. Karácsony 2022: 257, 259–260; Papp 2022: 147).

A deepfake sarokköve a szólás szempontjából az, hogy a deepfake maga is közlés (Mráz 2021: 252–261). A deepfake-készítő tehát közöl, a szólásszabadsághoz való jogát gyakorolja, amikor például a barátairól szellemes deepfake-videókat publikál (Van der Sloot – Wagensveld 2022: 3), de közöl akkor is, amikor hamis politikai üzeneteket manipulál deepfake-kel, vagy bosszúpornót készít expartnere arcképével (Mráz 2021: 252–253). Érdemes tehát hangsúlyozni, hogy a büntetőjogi (Ambrus 2021; Miskolczi–Szathmáry 2018: 140–141; G. Karácsony 2022: 310; Pantserev 2020: 50; Sorbán 2020: 85–100), az adatvédelmi és a polgári jogi (magánjogi) (Ebermann 2021: 39; Galyashina–Nikishin 2022: 4; Hine–Floridi 2022: 608–609) megközelítések mellett, a deepfake alapjogi aspektusa megkerülhetetlen a szabályozási irányok megértése szempontjából.

2. DEEPFAKE-VIDEÓ MINT VÉLEMÉNY?

Napóleon megjelenik történelemórán (Van der Sloot – Wagensveld 2022: 3), Lionel Messi futballsztár együtt dékázik 19 éves, 25 éves és 30 éves önmagával (W1), Salvador Dalíval csevegnek a látogatók a múzeumban (Veszelszki 2022: 33), David Beckham kilenc nyelven hívja fel a figyelmet a malária elleni küzdelemre (W2; Westerlund 2019: 41). A példák afféle premisszaként is szolgálhatnak a szólásként való definiáláshoz. Érdemes ugyanis leszögezni, hogy a deepfake sem mint szólás, sem mint technológia alapvetően és inherensen nem *rossz* és nem *tiltandó* (Pantserev 2020: 43–46). Ahogy Nicholas O'Donnell (2021: 720–724) fogalmaz, a deepfake-tartalmak jogi megközelítésének a szólásszabadság megértésén kell alapulnia. Természetesen léteznek válfajai, amelyek adatvédelmi aggályokat okoznak (Ambrus 2021), vagy büntetőjogi tényállásokat valósítanak meg (Btk. Kommentár 2022: 204. §, gyermekpornográfia), de nem lenne helyénvaló a deepfake teljes körű tiltása vagy nagymértékű korlátozása, hiszen azzal magát a szólást tiltanánk vagy korlátoznánk.

Azt állíthatjuk tehát, hogy a deepfake eszköz és nem tartalom, ezért nem az eszközt kell vizsgálni és korlátozás alá tenni, hanem a tartalmat kell egyedi esetekre vetítve megvizsgálni, és a jogi szabályozás alá vetni a jogellenességeket. A szólás kérdését érdemes egy provokatív példával bevezetni: tételezzük fel, hogy kikerül a közösségi médiára egy deepfake-videó, amelyen korunk egyik ismert színésze ül az M2-es metrón, és részegen ordibál. Mi lenne a helyes jogi értékelés, mennyiben minősül szólásnak a videó? A videó minden tekintetben hamis és dehonesztál: nagyfokú a dezinformációs értéke, hamis kontextuális elemek találhatóak benne, és célja a szereplő lejáratása (Galyashina–Nikishin 2022: 4). Mégis a videó mint közlés, függetlenül attól, hogy minden tekintetben hamis, nem feltétlenül vet fel súlyos, szólásszabadsággal kapcsolatos aggályokat. Felfogható-e esetleg a videó paródiaként (Mráz 2021: 258)? Lehet-e egy utalás vagy referencia egy filmes sze-

replésre? Értelmezhető-e szimbolikusan a tartalom? A továbbiakban a különböző jogalkotások és szabályozási kísérletek eltérő megoldásainak ismertetésével igyekszem megválaszolni e kérdéseket.

3. SZABÁLYOZÁSI MEGOLDÁSOK – PÉLDÁK AZ EGYESÜLT ÁLLAMOK ÉS EURÓPA KEZDEMÉNYEZÉSEI KÖZÜL, TOVÁBBI ALTERNATÍV JAVASLATOK

A szabályozási kísérletek megismerése előtt szükségszerű annak megértése, hogy egyáltalán miért kell bármiféleképpen szabályozni a deepfake-tartalmakat. Az új technológiák és az online kommunikáció szabályozásának kapcsán két versengő elv szül dilemmát. Egyrészt az új technológiák és kommunikációs csatornák nem „jogmentesek" (tehát az új technológiákat ugyanúgy kell szabályozni, mint bármely más, korábbi technológiát); minimumkövetelmény ezek kapcsán az alapvető jogok tiszteletben tartása (Abh1: [439]–[453]), másrészt viszont az is egyértelmű, hogy nem lehetséges pusztán a jog eszköztárára hagyatkozni, hogy az alapjogok gyakorlása és védelme garantált legyen (Klein 2018: 11). Ahogy Rob Cover fogalmaz (2022: 615): a deepfake elsősorban „társadalmi aggály". Ez a megjegyzés találó a deepfake jelenségre, hiszen a deepfake felhasználása több területen is negatív hatással járhat; lehet itt gondolni a közbeszédre és közvéleményre gyakorolt hatására, a személyiségi jogokra és magánélethez való jogra (Van der Sloot – Wagensveld 2022: 10), a politikai megtévesztésre vagy akár a bosszúpornóra.

A hamis információk terjedésében, ahogy Veszelszki fogalmaz, a deepfake a „nehézfegyverzet" (Veszelszki 2022: 33). A politikai dezinformáció és megtévesztés is valós veszély (Freelon–Wells 2020), érdemes ehelyütt felidézni Volodimir Zelenszkij ukrán elnök 2022. márciusi „beszédét", amelyben egy deepfake-felvételen arra kéri az ukrán katonákat, hogy hagyjanak fel a harccal, és térjenek haza (W3) vagy a „bódult Nancy Pelosinak", az amerikai Képviselőház demokrata elnökének hamisított videóját (Buo 2020: 2–3). Vaccari és Chadwick felmérésében (2020: 6) egy négy másodperces videó megtekintése során a nézők felét tévesztette meg a deepfake-technológia, a kísérletben részt vevők 35%-a bizonytalan volt, hogy a felvételen deepfake-tartalmat látott-e. Ez az arány azt is illusztrálja, hogy ha a 2020-as deepfake-technológia a tartalomfogyasztók felét megtévesztette, akkor az idő elteltével és a technológia fejlődésével ez az arány egyre aggasztóbb lehet – a hamis hírek elleni harc kontextusában különösen.

A deepfake-pornográfia még akutabb probléma, tekintve, hogy becslések szerint az összes deepfake-tartalom több mint 90%-a pornó vagy pornográf jellegű felvétel (Wang 2019; Sorbán 2020). A deepfake-pornográfia mint lehetséges szabályozási terület legégetőbb problémája nem maga a tartalom, hiszen a pornográf tartalmak élvezhetnek jogi védelmet (Sorbán 2020: 86–90), hanem az arckép nem

konszenzuális felhasználása (Gieseke 2020: 1488; Ellis 2018). Meskys és szerzőtársai (2020: 10–11) és Ambrus István (2021) is a büntetőjog eszközeit javasolja a védekezésre és a jogérvényesítésre. Végső soron a deepfake pszichológiai és etikai polémiák bölcsője is: hiszen egy személy arcával, hangjával, vonásaival, jellegzetességeivel gyakorlatilag bárki, bárhol, bárkinek elküldhet vagy gyárthat az adott egyénre nézve sértő, megalázó vagy káros tartalmat; nem ismert, nem belátható a deepfake valós kockázata.

3.1. Amerikai sodrásirányok – mit érnek a kezdeményezések harmónia nélkül?

Az amerikai szakirodalom leginkább a First Amendmentből (az amerikai alkotmány első kiegészítéséből) és a Supreme Court esetjogából vezeti le a deepfake kapcsolatát a szólásszabadsággal (O'Donnell 2021: 720; Sorbán 2020: 96). A First Amendment előírja és védi a szólásszabadságot, azzal, hogy a Kongresszus nem hozhat törvényt annak korlátozásával kapcsolatban (First Amendment 1791). A kérdés a deepfake kapcsán az, hogy van-e a manipulált, hamisított tartalomnak bármiféle fundamentális, sajátlagos értéke, vagy csak és kizárólag az „ártás" a célja a deepfake-tartalom közlésének (O'Donnell 2021: 721–724; Reid 2021: 211–217; New York Times vs. Sullivan 1964).

Ugyancsak fontos a deepfake a közlés által kiváltott közvetlen negatív hatás a szólásszabadság tekintetében, amely egy elbocsátástól kezdve anarchikus utcai lázadásokig minden hátrányos eseményt jelenthet (Kugler–Pace 2021: 666–668; O'Donnell 726; Brandenburg vs. Ohio 1969). Ez azt is jelenti, hogy egy deepfake-paródia élvezi a First Amendment védelmét, amennyiben paródiaként kategorizálja magát (Hall 2018: 62) – ez az elv többek között már a deepfake előtti esetjogokból is tisztán kiderül (Hustler vs. Falwell 1988: 52–53). Szintén korábbi esetjog támasztotta alá azt is, hogy a provokatív, sértő ábrázolást is védi a szólásszabadsághoz való jog védelme (Douglass vs. Hustler Magazine 1985), ami még nehezebbé teszi egy sértő deepfake-tartalom pontos megítélését (Gibson 2020: 269). E körben kiemelendő még a Communications Decency Act (CDA) 230. §-a, amely a „jó szamaritánus"-elvként elhíresült rendelkezésében védelmet nyújt a polgári jogi felelősséggel szemben az obszcén vagy sértő közlések jóhiszemű moderálása kapcsán a platformoknak (Gosztonyi 2022: 99–100; Papp 2022: 82–92). A CDA ugyanakkor inkább történelmi jelentőségű, hiszen a törvény 1996-os kihirdetése óta új technológiák garmadája jelent meg, és a szabály kikerülése meglehetősen egyszerű is a platformok részéről (Gieseke 2020: 1508–1509; Gosztonyi 2022: 101–102), ugyanakkor a CDA 230. § megújításával a deepfake tekintetében felvethető lenne a felelősség és a hatékonyabb eltávolítási mechanizmusok kérdése is (Gieseke 2020: 1509).

A deepfake kapcsán fontos röviden értekezni az eltérő tagállami gyakorlatokról. Kalifornia államban a deepfake elleni fellépést elsősorban politikai megfontolások indokolták: a tagállami választási törvényben a valamilyen politikai tisztséget betölteni kívánó jelölt a választást követő 60 napon felléphet az ellen a deepfake-tartalmat gyártó személlyel szemben, aki „le akarta járatni" a jelöltet a kampány során a tartalommal (Van der Sloot – Wagensveld 2022: 11; Californian Elections Code § 20010). A kaliforniai koncepció releváns problémákat kíván orvosolni (Ferraro 2019: 9–10), de hatásossága erősen megkérdőjelezhető: a nyomozást ugyanis nagyban megnehezíti az anonim deepfake-gyártók kilétének feltárása, ami végső soron aláássa a rendelkezés eredményességét (Cover 2022: 616). A kaliforniai szabályozás mintájára készült a texasi szabályozás is, amely elsőként tiltotta a politikai jelölteknek kárt vagy hátrányt okozó deepfake-videókat (Ferraro 2019: 13).

New York és Nebraska állam a közvetlen deepfake-tiltás felé fordult (Gibson 2020: 269, 282–283; Ferraro 2019: 12). A két tagállam a nyilvánosság védelme, illetve a „rossz szándékkal létrehozott" tartalmak készítését és közzétételét szankcionálja (Ferraro 2019: 12; S. 3805, 115th Cong. 2018: § 1041). A probléma ugyanaz viszont, mint a kaliforniai szabályozással: félő, hogy ezek a rendelkezések csupán szimbolikus jelentőségűek maradnak, mert a végrehajtásuk nem átlátható és nem előre látható, illetve nincs megbízható, kiforrott technológiai vagy jogi eszköz a deepfake-gyártók felkeresésére.

Virginia államban 2019. július 1. napján lépett hatályba a más személyről készült képek jogellenes terjesztéséről vagy értékesítéséről szóló törvény (Unlawful Dissemination or Sale of Images of Another Person; Ferraro 2019: 14; Va. Code Ann. § 18.2–386.2, HB 2678, SB 1736 2019). A törvény a nem konszenzuálisan „hamisan készített" képek és videók terjesztését vétségnek minősíti (HB 2678 VA 2019), az ilyen típusú cselekmény lehetséges szankciója pedig pénzbüntetéstől akár egy évig terjedő szabadságvesztésig is terjedhet.

Kiemelendő a tagállami kezdeményezéseken túl ívelő szabályozás, a Kongresszus 2019-es DEEP FAKES törvényjavaslata (a név maga is egy rövidítés: „Defending Each and Every Person from False Appearance by Keeping Exploitation Subject to Accountability Act"; G. Karácsony 2022: 311–312; Ferraro 2019: 4, 6), amely bár bizottsági szinten elbukott, mégis érdemleges lépés volt a deepfake illegális felhasználása ellen. A DEEP FAKES Act célja, hogy a lehető összes, deepfake által előrevetített jogellenes tevékenységet orvosolja (Ferraro 2019: 7). A törvényjavaslat szerint a deepfake-kel manipulált képek és videók megjelenítése során a készítők kötelesek vizuális figyelemfelhívást is közzétenni, amelyben jelzik, hogy az adott tartalom manipulált (DEEP FAKES Act 2018: § 1041). E kötelezettség elmulasztása súlyos, akár 150 ezer dolláros bírságot is maga után vonhat [DEEP FAKES Act 2018: §1041 f) 2.; G. Karácsony 2022: 311].

3.2. Európai szabályozási célpontok, magyar és kínai átszállással

Az európai szintű szabályozással kapcsolatban elsősorban az EU mesterséges intelligenciára vonatkozó rendelettervezetére (MI-rendelet) és a Digital Services Actre (DSA) érdemes kitérni. Az MI-rendelet szerint [2021: 52. cikk (3); ERPS 2021] a deepfake-rendszerek úgynevezett alacsony kockázattal járó mesterséges intelligenciát használó rendszerek (Fernandez 2022). Az MI-rendelet a jogellenes deepfake-tartalmakkal szembeni védekezés elsődleges módszereként a deepfake-tartalom megjelölését javasolja (G. Karácsony 2022: 312). Ennek értelmében a mesterségesintelligencia-rendszerek felhasználóinak a deepfake-technológiával manipulált tartalmak (képek, audió- vagy videótartalmak) generálásakor közlési kötelezettsége keletkezne, vagyis a tartalmat mesterségesen létrehozott vagy manipulált tartalomként kellene megjelölni [MI-rendelet 2021: 52. cikk (3)]. Heves kritika érte az MI-rendeletet, mivel azzal, hogy alacsony kockázati tényezővé minősítette a deepfake-technológiát, nagyban gyengült a hathatós és produktív végrehajtás lehetősége is (Fernandez 2022).

A DSA bár kevésbé szorosan kapcsolódik a deepfake-szabályozáshoz, mégis megemlítendő kiemelkedő jelentősége miatt az európai platformszabályozás terén. A 2022 októberében kihirdetett rendelet (OJL EU L 277: 2022) kötelezettségeket állapít meg a tárhelyszolgáltatókra, így a platformokra is, azzal a céllal, hogy a platformok szabályozásával a felhasználók és alapjogaik teljes körű védelme garantált legyen az online térben (Török 2022: 195–200). A DSA több újítással és megoldással igyekszik elérni e célokat; a platformokra extraterritoriális jelleggel is kiterjeszti az EU-s kötelezettségeket, és a bejelentési és cselekvési mechanizmusokkal magasabb szintű fellépést jelent a jogellenes tartalmakkal szemben (DSA 2022: 16. cikk; Gosztonyi 2022: 147). Ugyan számtalan kritika érte a DSA fogalomhasználatának tág jellegét vagy éppen a jogellenes tartalmakkal szembeni védekezés lehetséges elbürokratizálódását (Gosztonyi 2022: 148–150; Peukert 2021), a DSA kétséget kizárólag progresszív és pozitív lépés a platformszabályozás és ennek értelmében a jogellenes deepfake elleni fellépés terén.

Törvényi kerettel rendelkező európai megoldás a platformok aktív és felelős eljárásra kötelezése a jogellenes tartalmakkal, így akár a problémás deepfake-tartalmakkal szembeni fellépés is (vö. Gerstner 2020: 12–13). E tekintetben nemzeti szinten kiemelhető a német Netzwerkdurchsetzungsgesetz (NetzDG), amely a jogellenes tartalmakkal szemben a platformokra ró kötelezettségeket. A NetzDG rendelkezései alapján a törvény jogellenes tartalomnak tartja többek között a deepfake által leggyakrabban érintett tevékenységeket, így a német büntető törvénykönyvre, a Strafgesetzbuchra (StGB) hivatkozva, a gyermekpornográfia terjesztését, beszerzését és birtoklását (StGB 184b. §), a rágalmazást (StGB 187. §), a politikai életben szereplő személyekkel kapcsolatos hamis információközlést (StGB 188. §) vagy a magánélet megsértését fényképek vagy más képek készítésé-

vel (StGB 201a. §). A platformoknak, amennyiben ilyen tartalmat jelentenek nekik, olyan kötelezettsége is kialakul, hogy a német Szövetségi Bűnügyi Hivatalnak továbbítsák a felhasználó adatait, a jogellenesnek ítélt tartalom eltávolítása mellett (NetzDG 3a. §).

Érdemes megemlíteni a kommunikációs platformokról szóló osztrák törvényt (Kommunikationsplattformen-Gesetz, KoPl-G), amely a NetzDG-hez hasonlóan a platformokra nézve kötelezettségeket állapít meg a jogellenes tartalmakkal szembeni fellépés terén, különösen azzal, hogy a NetzDG-hez hasonló panaszkezelési és bejelentési rendszer létrehozását írja elő a platformoknak (KoPl-G 3–7. §§).

Bár 2022 előtt ki lehetett jelenteni, hogy az angolszász szabályozás elmarad a német és osztrák kezdeményezésektől (Farish 2020: 48), Skóciában már fellelhető valamiféle deepfake-szabályozás. A skót jog egyik passzusa szerint az intim fényképek és filmek nem konszenzusos nyilvánosságra hozatala bűncselekménynek minősül, függetlenül annak jellegétől, így a manipulált felvételek is a törvény hatálya alá tartoznak (Jones–Jones 2022: 21; Abusive Behaviour and Sexual Harm Act 2016: pt1. 2.). Az Egyesült Királyság Igazságügyi Minisztériuma 2022 végén az Online Safety Bill módosítását javasolta; a módosítás szerint a biometrikus módszerekkel manipulált intim képek megosztása engedély nélkül kriminalizált cselekmény lesz (Mascellino 2022). Egyelőre a törvénymódosítás gyakorlati hatásai ismeretlenek, mindenesetre az első reakciók többségében, még ha józan visszafogottsággal is, de optimisták (Dawson 2022; Hern 2022).

Röviden kitérve a magyar deepfake-perspektívákra is: egyelőre semmiféle specifikus vagy partikuláris szabályozás nem vonatkozik a deepfake használatával elkövetett bűncselekményekre (így különösen a deepfake-pornográfiára) vagy egyéb törvényellenes tevékenységekre (Sorbán 2020: 99–100). Ennek ellenére a magyar szabályozási környezet, kiemelten a 2012. C. törvény a Büntető Törvénykönyv (Btk.) releváns rendelkezései, például a zaklatás, rágalmazás, becsületsértés, szexuális zsarolás (Sorbán 2020: 99), alkalmazhatóak lehetnek deepfake kép- és hangfelvételekre. Mezei Kitti és Szentgáli-Tóth Boldizsár (2022: 248) felveti a nem szigorúan individuumhoz kötött bűncselekmények kapcsán is a deepfake-szabályozás kérdését: a szerzőpáros szerint a deepfake-tartalmak alkalmasak lehetnek akár a Covid-19 pandémiával kapcsolatos álhírek terjesztésére is. A szerzőpáros által felvetett Covid-dezinformációra ugyanakkor létezik magyar szabályozás; bár az új rémhírterjesztésre vonatkozó bekezdés [Btk. 337. § (2)] vegyes reakciókat váltott ki a magyar jogirodalomban (vö. Bencze–Győry 2021; Domokos 2021; Drinóczi 2021; Lendvai 2023), a rendelkezés alkalmazható lehet a deepfake-technológiát alkalmazó rémhírterjesztés megállítására és szankcionálására. A deepfake szabályozási igénye kapcsán Ambrus (2021) csatlakozik Sorbánhoz: véleménye szerint nem szükséges önálló tényállásba foglalni a deepfake-et, mivel „a deepfake kategória nem hordoz plusz társadalomra veszélyességet", míg Miskolczi és Szathmáry (2018: 140–141) szerint szükséges lenne magasabb szintű jogi

védelmet alkotni a deepfake kapcsán, különösen a deepfake által okozott adatvédelmi problémák kiszűrése miatt.

Kína szabályozása kapcsán röviden érdemes kitérni arra, hogy milyen alternatív, kevésbé óvatos deepfake-szabályozások léteznek. 2022. január 28-án a Kínai Kibertér Hatóság kiadott egy, a deepfake-szintézist használó internetes információs szolgáltatások adminisztrációjáról értekező rendelkezéscsomagot (Creemers–Webster 2022). A deep-szintézis technológiát a szabályozás ernyőfogalomként használja (2. §), amely lefedi gyakorlatilag az összes audiovizuális tartalmat (Hine–Floridi 2022: 608). A szabályozás célja a deepfake elleni konkrét, komoly és szigorú fellépés és eltérés az amerikai vagy európai szabályozásoktól, hogy a kínai deepfake-rendelet egyedülálló módon a deepfake-szintézis programokat, szoftvereket szolgáltatókat célozza, nem a platformokat vagy a deepfake-szoftverek felhasználóit (Hine–Floridi 2022: 608–609). Ez az újfajta jogi gondolkodás összességében pozitív visszajelzéseket kapott, tekintve, hogy előreláthatólag nagyobb védelmet élvezhetnek majd az internethasználók, és szigorúbb, jogilag biztosított elszámoltathatósággal kell figyelembe venniük a deepfake-technológiát fejlesztő cégeknek (Hine–Floridi 2022: 610; Cai–Zhang 2022; Yan 2022).

3.3. „A felesleges törvények gyengítik a szükségeseket" (Montesquieu) – biztosan a jog a legjobb szabályozási megoldás?

Az első deepfake-szabályozó eszköz nem törvény vagy valamiféle esetjogi döntés kapcsán alakult ki: a Reddit nevű platform 2018-ban úgy döntött, hogy korlátozni fogja a deepfake-tartalmakat az oldalon (Van der Sloot – Wegensveld 2022: 3). A szabályozás előzménye a deepfake „subredditen" (gyakorlatilag „alblogon", ahol specifikus témakörök szolgálják az alcsoport beszéd- és vitatémáit) terjedő deepfake-pornográf tartalmak elszaporodása volt (Van der Sloot – Wagensveld 2022: 3; Sorbán 2020: 90).

A deepfake-gyártás mellett aktívan fejlődnek a deepfake-felismerő szoftverek is (G. Karácsony 2022: 313–314). Ilyen technikai irány a többfaktoros azonosítási rendszerek implementációja vagy a hang- és képalapú azonosítása, bár ezeknek a biztonsági megoldásoknak a kijátszása továbbra sem okoz különösebb gondot a fejlettebb deepfake-szoftvereknek (G. Karácsony 2022: 306). A kizárólag technológiai nóvumokra hagyatkozó szabályozási módszer másik hátulütője, hogy jelenleg – bár egyre nagyobb szabású kutatások zajlanak – egyértelműen kijelenthető, hogy nem létezik és a közeljövőben nem is valószínű, hogy létrejön 100%-os biztonsággal, abszolút precizitással dolgozó deepfake-felismerő rendszer (Pantserev 2020: 48). A technológiai eszköztár emberi erőforrással való kiegészítése, bár hatékonyabb javaslatnak tűnhet, mégis magas költségekkel és ismételten nem megfelelő hatékonyságú védelemmel járna a deepfake-tartalmakkal szemben (Cover 2022: 617). Van

der Sloot és Wagensveld (2022: 12) amellett érvel, hogy a deepfake szabályozása a jog eszköztárával ingoványos terület, nem a törvényhozás, jóval inkább a szabályok végrehajtása szempontjából. Ahogy a szerzőpár fogalmaz, „a deepfake demokratizálódik"; vagyis már nemcsak a nagy filmstúdiók engedhetik meg maguknak, hogy „feltámasszák" a múlt hírességeit a vásznon a technológia segítségével, hanem gyakorlatilag bármely felhasználó is képes deepfake-tartalmat gyártani, legtöbbször ingyenesen hozzáférhető alkalmazások segítségével (Cover 2022: 613).

A jelen fejezet a deepfake-re egy trianguláris szabályozási modellt javasol. A modell váza az említett szabályozási metódusok és filozófiák elegyítése lenne, magyarán az állami és nemzetközi törvényhozás harmonizált kapcsolódása a platformok saját ellenőrző tevékenységével, felhasználva és fejlesztve a folyamatosan fejlődő, a deepfake felismerésére és azonosítására irányuló technológiai lehetőségeket. Ez a háromlépcsős modell természetesen számtalan kihívással szembesülne, így a szabályok harmonizálása, a törvényi szabályozás összeegyeztetése a platformok privát szabályozásával (ehelyütt például megoldás lehet, ha a platformok egy ellenőrző bizottság segítségét vennék igénybe szabályaik kialakítása során; Hall 2018: 73), a fogalmi keretek egységesítése kapcsán (erre Mathilde Pavis alapos és jól definiált rendszert javasolt; Pavis 2021: 981–982), nem is beszélve a gazdasági érdekek figyelembevételéről. Ugyanakkor a közös, megosztott fellépés átfogó, holisztikus megoldást jelenthetne (Meskys et al. 2020: 11–12) a felhasználók és egyének sérülékenységének mérséklése (Cover 2022: 617), a kommunikációs csatornák fejlődésének megőrzése és a szólásszabadság védelme érdekében (Graber-Mitchell 2021; Mráz 2021: 262–263). A hármas szabályozás ugyanakkor valós változást csak akkor tudna elérni, ha a felhasználók tudatosítása és figyelemfelhívása kiemelt szerepet játszana a deepfake-technológia valamennyi lényeges, egyént érintő vonatkozásában (Hall 2018: 74; Farish 2020: 48).

4. DERMESZTŐ HATÁS

A közlés és legfőképpen annak korlátozásának arányossága (vagy éppen aránytalansága) kapcsán szót kell ejteni a „chilling effect"-ről, azaz a dermesztő hatásról. A chilling effect alapvetően európai jogi koncepció (Pech 2021: 2), bár pontos, kiforrott definíciója nincs (EU Bizottság Rule of Law Report 2020). Ha mégis kísérletet tennénk a fogalom meghatározására, akkor a dermesztő hatás egy olyan negatív következménnyel járó jogi eljárás vagy intézkedés az állam részéről, amely visszatartja az eljárás vagy intézkedés címzettjeit attól, hogy jogaikat (lásd a szólásszabadság jogát) gyakorolják vagy kötelezettségeiket teljesítsék. A „dermesztés" hatása pedig abban a félelemben nyilvánul meg, amelyet a címzettek kifejeznek, tartva attól, hogyha mégis gyakorolnák jogaikat vagy kötelezettségeiket, valamiféle szankció vagy eljárás alanyai lennének (Pech 2021: 4).

Bár a szólásszabadság joga nem abszolút, értve ezen, hogy lehet korlátozni bizonyos esetekben, amennyiben azt törvény írja elő azt, és szükséges a korlátozás (ICCPR 1966: §19; Baumbach 2018: 93–96). A deepfake kontextusában ilyen dermesztő hatással járhat egy túlságosan is szigorú, tág szabályozás (Mráz 2021: 260–261). Tételezzük fel, hogy a korábban említett, híres színészt érintő deepfake-videó kapcsán törvény születne a deepfake teljes tiltásáról, a deepfake gyártása és publikálása pedig büntetőjogi szankciókat vonna maga után. Az új törvénykezés kétséget kizáróan megbénítaná a közlés módját, és nagymértékében, aránytalanul avatkozna bele az egyének szólásszabadsághoz való jogába (Penney 2022: 103–105).

A dermesztő hatás és a deepfake kapcsolatára magyar példa is van; a 4K! – Negyedik Köztársaság párt kampányvideója (Mráz 2021: 258). Az Alkotmánybíróság elé került ügy szerint egy kampányvideóban „egy katonai egyenruhába bújt majomnak öltözött ember szerepel, aki korábbi magyar miniszterelnökök hangjára tátog", és ennek a közlését egy médiaszolgáltató megtagadta (ABh2 2014: [1]–[2]). Az Alkotmánybíróság elutasította a kérelmet azzal, hogy a felvétel megtagadása jogszerű volt a médiaszolgáltató részéről, hivatkozva az Alaptörvény II. cikkére és IX. cikk (4) bekezdésére, amelyek a véleménynyilvánítás szabadságának korlátozásáról szól az emberi méltóság megóvásának érdekében. A döntés ellentmondásos (Mráz 2021: 258), többek között a vonatkozó Európai Emberi Jogok Bírósága (a továbbiakban: EJEB) döntések fényében is. Az EJEB ugyanis több ízben kiemelte, hogy egyrészt a közszereplőknek – így különösen politikai szereplőknek (ECtHR5 2014: 67. pont; ECtHR6 2012: 45. pont) – nagyobb tűrési kötelezettségük van a provokatív és adott esetben sértő szólással szemben (ECtHR1 2017: 54–55. pontok; ECtHR2 2017: 35. pont; ECtHR3 1986: 42. pont; ECtHR4 2016: 42. pont). Másrészt bár a nagyobb tűrési kötelezettség nem terjed ki az indokolatlanul bántó és személyeskedő közlésekre (ECtHR7 2021: 55. pont), a politikai vita és a közbeszéd tárgyát képező témák, mint például a választások és a kampányidőszak során ugyancsak megengedett a provokatív, túlzó, kritikus szólás (ECtHR8 2005: 46. pont). A deepfake-tartalmak kapcsán többször előkerül a hamis tényállítás kérdése is (Mráz 2021: 260–262; Sorbán 2020: 90–91). Ugyan a hamis beszéd alacsonyabb szintű alkotmányos védelmet élvez (Gibson 2020: 285–286), mégis megemlítendő a deepfake-tartalmak „művészi" közlés jellege. E tekintetben tehát előremutató lenne egy kevésbé kategorikus jogi perspektíva, amely a deepfake-tartalmat, kontextustól függően (Gerstner 2020: 5; Hustler vs. Falwell 1988), nem csak és kizárólag a tényállítás–véleménynyilvánítás rendszerében képzeli el. Példának okáért, Nicolas Cage amerikai színész szatirikus rámontírozása bizonyos felvételekre inkább az irónia mint kommunikációs eszköz kivetülése, mintsem hamis tényállítás (Van der Sloot – Wagensveld 2022: 3). A hamis tényállítások kérdése azért is nehézkes továbbá, mert megállapításuk igencsak kontroverziális és sok esetben nehézkes is, különösen annak fényében, hogy a szólásszabadság

védelme kiterjed a sokkoló, bántó tartalmakra (ECtHR9 2005; ECtHR10 2010: 137. pont; ECtHR11 2019: 158. pont) és az ironikus és szatirikus közlésekre is (ECtHR12 2013: 29. pont; ECtHR13 2013: 60. pont; ECtHR14 2009: 27. pont).

A deepfake, a jog és a szólásszabadság kapcsolatának gyakorlatban való megjelenése kapcsán érdemes lehet egy esetjogi analógiát vonni. Szatirikus tartalom, egy képzeletbeli interjú volt a tárgya Nikowitz and Verlagsgruppe News GmbH v. Austria (2007) esetnek („Nikowitz-eset"). Egy osztrák magazin újságírója egy ironikus, szatíraszerű esszét publikált, amelyben bírálta az osztrák lakosság és média reakcióját a közúti balesetek kapcsán, mivel Hermann Maier, akkori osztrák sífutóbajnok is közúti szerencsétlenség során sérült meg (Nikowitz-eset 2007: 4–6. pontok). A cikkben egy fiktív interjú passzusát is közölte a lap; Stefan Eberhart sífutó, Maier egyik legnagyobb riválisa, a lap közlésében úgy reagált sporttársa sérülésére, hogy „remek, végre én is nyerek valamit, és remélem, hogy az a rohadt kutya elcsúszik a mankójával, és a másik lába is eltörik" (Nikowitz-eset 2007: 7. pont). Eberhart becsületsértés miatt eljárást indított a lap ellen (Nikowitz-eset 2007: 8. pont). Az illetékes bécsi bíróságok első és másodfokon is Eberhart javára ítéltek, és a lapot sérelemdíj megfizetésére kötelezték. Érdekesség, hogy a Bécsi Fellebbviteli Bíróság több ízben is kifejtette az ítéletében, hogy az ilyen típusú közlések, mint az ügyben érintett cikk, értelmezése nagyon magas szintű intelligenciát és koncentrációt igényelnek (Nikowitz-eset 2007: 11. pont). Az ügy az Emberi Jogok Európai Bírósága elé került. A bíróság megállapította, hogy szatíra esetén is különbséget kell tenni az értékítéletek és tényállítások között, és arra a következtetésre jutott, hogy a képzelt interjú nem tényállítás volt, hanem szatírába bújtatott értékítélet, magyarán humornak szánt véleménynyilvánítás.

A dermesztő hatás szempontjából tehát nulladik pontként annak körvonalazása lenne célszerű, hogy egyáltalán mikor és milyen körülmények között tilos a deepfake használata, akár szatíraként vagy ironikus tartalmaknál (Pantserev 2020: 50). Ebből a példából ugyanakkor jól érzékelhető, mennyire komplex és alapos háttértudást igényel egy fiktív, „hamisított" tartalom megítélése – nem beszélve arról, hogy a deepfake esetében sokkal árnyaltabb és sokkal részletesebb kontextuális tényállási elemek vizsgálata is szükséges, különösen a deepfake-tartalom által kiváltott hatások tekintetében (Meskys et al. 2020: 24–31).

5. KÖVETKEZTETÉSEK

A kérdésre tehát, hogy el kell-e távolítani az írás elején említett közkedvelt színész videóját, bizonytalan választ kapunk. A két vizsgált kérdéskör előkérdése azonos: lehetséges-e a deepfake-videókat hamis audiovizuális tényállásoknak értelmezni? A deepfake és a szólásszabadság kapcsolata a dermesztő hatás viszonylatában igen komplex jogi szabályozást és megfontolást igényel. A téma érzékenységét még in-

kább fokozza, hogy körültekintő és beható vizsgálatra lenne szükséges minden egyedi esetben, hogy az adott videó közzététele és a sérelem között milyen okozati vagy tartalmi kapcsolat van (Van der Sloot – Wagensveld 2022: 11). Ugyan egyes szakértők úgy vélik, hogy a deepfake-készítéssel járó potenciális veszélyek jóval nagyobb kockázatot jelentenek, mint a deepfake tiltásának esetleges dermesztő hatása a szabályozásnak a szólásszabadságra (Pesetski 2021: 504), elengedhetetlennek tartom a szabályozás premisszájává tenni olyan pragmatikus szabályozási attitűdöt, amely a szabályozás során előtérbe helyezné az empirikus hatásvizsgálatok eredményét (Mráz 2021: 261), és egyben ügyelne arra is, hogy az új kommunikációs platformként szolgáló technológiai eszközök kialakítását és fejlesztését ne bénítsa meg (lásd LibertyWorks Inc vs. Commonwealth 2021: HCA 18, [95]). E feltételek figyelembevételével a deepfake-szabályozás a lehető legkisebb mértékben vethetné fel a dermesztő hatás által okozott polémiákat, különösen a politikai beszéd témakörében (Ray 2021: 1007).

SZAKIRODALOM

Ambrus István 2021: *Digitalizáció és büntetőjog.* Budapest: Wolters Kluwer. Új Jogtár verzió.

Baumbach, Trine 2018: Chilling Effect as a European Court of Human Rights' Concept in Media Law Cases. *Bergen Journal of Criminal Law and Criminal Justice*, 6/1: 92–114.

Bencze Mátyás – Győry Csaba 2021: Hírek szárnyán: a rémhírterjesztés bűncselekménye és a jogbiztonság. *Magyar Tudomány*, 182/5: 614–624.

Buo, Shadrack Awah 2020: The Emerging Threats of Deepfake Attacks and Countermeasures. Cornell University. *arXiv*, december 14.

Cai, Ron – Zhang, Andy 2022: China's Draft Legislation on Deep Synthesis Technologies. *Zhong Lun*, március 29. https://www.zhonglun.com/Content/2022/03-29/1450181043.html [2022. 12. 14.]

Citron, Danielle K. – Chesney Robert 2019: *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security.* Boston: Boston University School of Law.

Cover, Rob 2022: Deepfake culture: the emergence of audio-video deception as an object of social anxiety and regulation. *Continuum Journal of Media & Cultural Studies*, 36/4: 609–621.

Creemers, Rogier – Webster, Graham 2022: Translation: Internet Information Service Deep Synthesis Management Provisions (Draft for Comment). *Stanford University – Digichina*, február 04. https://digichina.stanford.edu/work/translation-internet-information-service-deep-synthesis-management-provisions-draft-for-comment-jan-2022/ [2022. 12. 14.]

Domokos Andrea 2021: Egyes, a járványhoz kapcsolódó büntetőjogi szabályok különleges jogrend idején. *Glossa Iuridica*, VII (különszám: Jog és vírus): 73–83.

Drinóczi Tímea 2021: Az Alkotmánybíróság határozata a különleges jogrend idején megvalósuló rémhírterjesztésről. *Jogesetek Magyarázata*, 1: 3–13.

Ebermann, Axel 2021: The Effects of Deepfakes and Synthetic Media on Communication Professionals. *CUNY Academic Works*, ősz: 1–155.

Farish, Kelsey 2020: Do deepfakes pose a golden opportunity? Considering whether English law should adopt California's publicity right in the age of the deepfake. *Journal of Intellectual Property Law & Practice*, 15/1: 40–48.

Fernandez, Angelica 2022: Regulating Deep Fakes in the Proposed AI Act, *Medialaws*, március 23. https://www.medialaws.eu/regulating-deep-fakes-in-the-proposed-ai-act/ [2022. 12. 14.]

Ferraro, Matthew F. 2019: Deepfake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media. WilmerHale report: Deepfake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media. *WilmerHale*, szeptember. 25. https://www.wilmerhale.com/en/insights/client-alerts/20190925-deepfake-legislation-a-nationwide-survey [2023.02.09.]

Freelon, Deen – Wells, Chris 2020: Disinformation as Political Communication. *Political Communication*, 37: 145–156.

G. Karácsony Gergely 2022: „Ideje megváltoztatni" – Az arcfelismerő rendszerek alkalmazásának alapjogi kockázatai a közösségi média és a deepfake korában. In: Török Bernát – Ződi Zsolt (szerk.): *Az internetes platformok kora.* Budapest: Ludovika Egyetemi Kiadó. 299–318.

Galyashina, Elena Igorevna – Nikishin, Vladimir 2022: The protection of megascience projects from deepfake technologies threats: information law aspects. *Journal of Physics: Conference Series*, 2210: 1–8.

Gerstner, Erik 2020: Face/Off: "DeepFake" Face Swaps and Privacy Laws. *Defense Counsel Journal*, 1–14.

Gibson, Kareem 2020: Deepfakes and Involuntary Pornography: Can Our Current Legal Framework Address This Technology? *Wayne Law Review*, 66: 259–289.

Gieseke, Anne P. 2020: „The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography. *Vanderbilt Law Review*, 75/5: 1479–1515.

Gosztonyi Gergely 2022: *Cenzúra Arisztotelésztől a Facebookig.* Budapest: Gondolat Kiadó.

Graber-Mitchell, Nicolas 2021: Artificial Illusions: Deepfakes as Speech. *Intersect*, 14/3: 1–19.

Hall, Holly Kathleen 2018: Deepfake Videos: When Seeing Isn't Believing. *Catholic University Journal of Law and Technology*, 27/1: 51–76.

Hine, Emmie – Floridi, Luciano 2022: New deepfake regulations in China are a tool for social stability, but at what cost? *Nature Machine Intelligence*, 4: 608–610.

Jones, Karl – Jones, Bethan 2022: How robust is the United Kingdom justice system against the advance of deepfake audio and video? *InfoTech 2022 IEEE International Conference on Information Technologies*, 9: 13–24.

Karsai Krisztina 2022: *Nagykommentár a Büntető Törvénykönyvhöz* [Btk. Kommentár]. Új Jogtár verzió. Budapest: Wolters Kluwer.

Klein Tamás 2018: Az online diskurzusok egyes szabályozási kérdései. In: Klein Tamás (szerk.): *Tanulmányok a technológia- és cyberjog néhány aktuális kérdéséről.* Budapest: NMHH Médiatudományi Intézet. 11–39.

Kugler, Matthew B. – Pace, Carly 2021: Deepfake Privacy: Attitudes and Regulation. *Northwestern University Law Review*, 116/3: 611–680.

Lendvai, Gergely Ferenc 2023: „Of Covid, [say] nothing but the truth": Reflections on the consequences on media platforms and freedom of expression principles of the new scaremongering rules implemented in the Hungarian Penal Code during the pandemics. In: Gosztonyi, Gergely – Lazar, Elena (szerk.): *Media Regulation during the COVID-19 Pandemic: A Study from Central and Eastern Europe*. Cambridge: Ethics Press. (kézirat)

Lyu, Siwei 2020: Deepfake Detection: Current Challenges and Next Steps. *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*: 1–6.

Meskys, Edvinas – Kalpokiene, Julija – Jurcys, Paulius 2020: Regulating deep fakes: legal and ethical considerations. *Journal of Intellectual Property Law & Practice*, 1: 1–13.

Mezei Kitti – Szentgáli-Tóth Boldizsár 2022: Az online platformok használatában rejlő veszélyek: a dezinformáció és a kibertámadások jogi kockázatai. In: Chronowski Nóra – Szentgáli-Tóth Boldizsár – Szilágyi Emese (szerk.): *Demokrácia-dilemmák. Alkotmányjogi elemzések a demokráciaelv értelmezéséről az Európai Unióban és Magyarországon*. Budapest: ELTE Eötvös Kiadó. 241–262.

Mráz Attila 2021: Deepfake, demokrácia, kampány, szólásszabadság. In: Török Bernát – Ződi Zsolt (szerk.): *A mesterséges intelligencia szabályozási kihívásai*. Budapest: Ludovika Egyetemi Kiadó. 249–277.

O'Donnell, Nicholas 2021: Have We No Decency? Section 230 and the Liability of Social Media Companies for Deepfake Videos. *University of Illinois Law Review*, 2: 701–740.

Pantserev, Konstantin A. 2020: The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability. In: Jahankhani, Hamid – Kendzierskyj, Stefan – Chelvachandran, Nishan – Ibarra, Jaime (szerk.): *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*. Cham: Springer Switzerland. 37–56.

Papp János Tamás 2022: *A közösségi média szabályozása a demokratikus nyilvánosság védelmében*. Budapest: Wolters Kluwer.

Paris, Britt – Donovan, Joan 2019: Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence. *Datasociety*, szeptember 18. https://datasociety.net/library/deepfakes-and-cheap-fakes/ [2022. 12. 14.]

Pavis, Mathilde 2021: Rebalancing our regulatory response to Deepfakes with performers' rights. *The International Journal of Research into New Media Technologies*, 27/4: 974–998.

Pech, Laurent 2021: *The Concept of Chilling Effect*. Open Society Foundations.

Penney, Jonathon W. 2022: Understanding Chiling Effect. *Minnesota Law Review*, 106: 1451–1530.

Pesetski, Anna 2021: Deepfakes: A New Content Category for a Digital Age. *William & Mary Bill of Rights Journal*, 29/2: 503–532.

Ray, Andrew 2021: Disinformation, Deepfakes and Democracies: The Need for Legislative Reform. *UNSW Law Journal Volume*, 44/3: 983–1013.

Reid, Shannon 2021: The Deepfake Dilemma: Reconciling Privacy and First Amendment Protections. *Journal of Constitutional Law*, 23: 209–238.

Sorbán Kinga 2020: A bosszúpornó és deepfake-pornográfia büntetőjogi fenyegetettségének szükségességéről. *Belügyi Szemle*, 10: 81–104.

Tariq, Shahroz – Lee, Sangyup – Woo, Simon S. 2021: One Detector to Rule Them All: Towards a General Deepfake Attack Detection Framework. *WWW '21: Proceedings of the Web Conference April*, 3625–3637.

Török Bernát 2022: A szólásszabadság a közösségi platformokon és a Digital Services Act. In: Török Bernát – Ződi Zsolt (szerk.): *Az internetes platformok kora.* Budapest: Ludovika Egyetemi Kiadó. 195–208.

Vaccari, Cristian – Chadwick, Andrew 2020: Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 1–13.

Van der Sloot, Bart – Wagensveld, Yvette 2022: Deepfakes: regulatory challenges for the synthetic society. *Computer Law & Security Review*, 46/105716: 1–15.

Veszelszki Ágnes 2021: deepFAKEnews: Az információmanipuláció új módszerei. In: Balázs László (szerk.): *Digitális kommunikáció és tudatosság.* Budapest: Hungarovox Kiadó. 93–105.

Veszelszki Ágnes 2022: A tudományos influencerektől a deepfake-ig. A legújabb tudománykommunikációs lehetőségek. *Filológia.hu*, 13/1–4: 27–39.

Westerlund, Mika 2019: The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9/11: 39–52.

FORRÁSOK ÉS JOGI DOKUMENTUMOK

Dawson, Brit 2022: Will the Online Safety Bill really stop deepfake porn? *Dazed*, december 9. https://www.dazeddigital.com/life-culture/article/57748/1/will-the-online-safety-bill-stop-deepfake-porn-sex-worker [2022. 12. 14.]

Ellis, Emma Grey 2018: People Can Put Your Face on Porn—and the Law Can't Help You. *Wired*, január 26. https://www.wired.com/story/face-swap-porn-legal-limbo/ [2022. 12. 14.]

European Parliamentary Research Service 2021: Tackling deepfakes in European policy. *European Parliament*, július 31. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2021)690039 [2023. 02. 09.]

Hern, Alex 2022: Online safety bill will criminalise 'downblousing' and 'deepfake' porn. *The Guardian*, november 24. https://www.theguardian.com/technology/2022/nov/24/online-safety-bill-to-return-to-parliament-next-month [2022. 12. 14.]

Mascellino, Alessandro 2022: UK unveils new laws to protect 'revenge porn' victims, including from deepfakes. *Biometric Update*, november 29. https://www.biometricupdate.com/202211/uk-unveils-new-laws-to-protect-revenge-porn-victims-including-from-deepfakes [2022. 12. 14.]

Peukert, Alexander 2021: Five Reasons to be Skeptical About the DSA. *Verfassungsblog*, augusztus 31. https://verfassungsblog.de/power-dsa-dma-04/ [2022. 12. 14.]

Wang, Chenxi 2019: Deepfakes, Revenge Porn, And The Impact On Women. *Forbes*, november 1. https://www.forbes.com/sites/chenxiwang/2019/11/01/deepfakes-revenge-porn-and-the-impact-on-women/?sh=11d463c71f53 [2022. 12. 14.]

W1 = Adidas Football (YouTube) 2022: The Impossible Rondo. *YouTube*, 11. 18. https://www.youtube.com/watch?v=_h6aWfAhiis [2022. 12. 14.]

W2 = Zero Malaria Britain (YouTube) 2019: David Beckham speaks nine languages to launch Malaria Must Die Voice Petition. *YouTube*, 04. 19. https://www.youtube.com/watch?v=QiiSAvKJIHo [2022. 12. 14.]

W3 = *Deepfake presidents used in Russia-Ukraine war. BBC*, 2022. 03. 18. https://www.bbc.com/news/technology-60780142 [2022. 12. 14.]
Yian, Li 2022: Using ‚deepfake' technology in China may require identity verification in future: cyberspace regulator. *ECNS*, 01. 29. http://www.ecns.cn/news/sci-tech/2022-01-29/detail-ihavfwhh0343801.shtml [2022. 12. 14.]

Nemzetközi jogi dokumentumok

International Covenant on Civil and Political Rights, 1966 (ICCPR)
Az Európai Parlament és a Tanács Rendelete: A Mesterséges Intelligenciára Vonatkozó Harmonizált Szabályok (a Mesterséges Intelligenciáról Szóló Jogszabály) Megállapításáról és Egyes Uniós Jogalkotási Aktusok Módosításáról [COM(2021) 206; 2021/0106(COD)]
Az Európai Unió Hivatalos Lapja, L 277, 2022. október 27. (OJL EU L 277: 2022)
Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)

Amerikai jogi dokumentumok és esetek

Constitution of the United States: First Amendment 1791
Californian Elections Code
Communications Decency Act of 1996 (1996)
H.R.2395 – DEEP FAKES Accountability Act (DEEP FAKES Act)
S. 3805, 115th Cong. 2018
Va. Code Ann. § 18.2–386.2, HB 2678, SB 1736 - 2019
Brandenburg vs. Ohio, 395 U.S. 444 (1969)
Douglass vs. Hustler Magazine, Inc., 607 F. Supp. 816 (N.D. Ill. 1985)
Hustler Magazine, Inc. vs. Falwell, 485 U.S. 46 (1988)
New York Times Company vs. Sullivan, 376 US 254 (1964)

Európai nemzeti jogi dokumentumok és esetek

Magyarország Alaptörvénye
2012. évi C. törvény a Büntető Törvénykönyvről
Abusive Behaviour and Sexual Harm (Scotland) Act 2016
Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG)
Kommunikationsplattformen-Gesetz – KoPl-G
Strafgesetzbuch für das Deutsche Reich (StGB)
LibertyWorks Inc v Commonwealth of Australia [2021] HCA 18
Abh1: 19/2014. (V. 30.) Alkotmánybírósági határozat
Abh2: 3122/2014. (IV. 24.) Alkotmánybírósági határozat

Nikowitz and Verlagsgruppe News GmbH v. Austria App no. 5266/03ECtHR1: Kącki vs. Poland App no. 10947/11
ECtHR2: Kapsis and Danikas v Greece App no. 52137/12
ECtHR3: Lingens vs. Austria App no. 9815/82
ECtHR4: Nadtoka vs. Russia App no. 38010/05
ECtHR5: Axel Springer AG vs. Germany (No.2) App no. 48311/10
ECtHR6: Tuşalp vs. Turkey App. nos 32131/08; 41617/08
ECtHR7: Dickinson vs. Turkey App. no 25200/11
ECtHR8: Paturel vs. France App no. 54968/00
ECtHR9: Vérités Santé Pratique Sarl vs. France App no. 74766/01
ECtHR10: Dink vs. Turkey App nos. 2668/07, 6102/08, 30079/08, 7072/09 and 7124/09
ECtHR11: Khadija Ismayilova vs. Azerbaijan App nos. 65286/13; 57270/14
ECtHR12: Welsh et Silva Canha vs. Portugal App no. 58106/15
ECtHR13: Eon vs. France App no. 26118/10
ECtHR14: Alves Da Silva vs. Portugal App no. 41665/07

ESZTERI DÁNIEL

A deepfake-technológia adatvédelmi értékelése a GDPR tükrében

A fejezet célja a deepfake-technológia értékelése a személyes adatok védelméhez fűződő alapvető jog érvényesülése szempontjából. A technológia alkalmazásának adatvédelmi problémáit az Európai Unió általános adatvédelmi rendeletének (GDPR) vonatkozó előírásai alapján elemzem. A fejezet a személyiséglopás és ezzel összefüggésben a valóság manipulálásának történeti előzményeivel kezdődik, majd a deepfake-technológia hátterét abból a szempontból mutatja be, hogy a folyamat egyes részelemei (szoftver tanítása, a deepfake elkészítése, majd annak felhasználása) milyen személyes adatkezelési műveletekkel járnak együtt. E mozzanatok kapcsán az adatkezelői felelősség is szóba kerül. A fejezet a jelenlegi jogérvényesítési lehetőségek és a legújabb, a jelenséget kezelni hivatott jogalkotási irányok felvázolásával zárul. Megállapíthatjuk, hogy a technológia használata kapcsán az érintettek magánszférájára jelentett kockázatok jelenleg nem kezelhetőek hatékonyan a legtöbb esetben, a tartalmakat előállító adatkezelők rendkívül nehéz beazonosíthatósága miatt. Az EU-s jogalkotás jelenlegi irányai azonban előremutató intézkedéseket tartalmaznak, amelyek gyakorlati végrehajthatósága azonban szintén kihívásokat hordoz magában.

Kulcsszavak: személyes adat, adatvédelem, GDPR, gépi tanulás, mesterséges intelligencia, személyiséglopás

1. BEVEZETÉS: A SZEMÉLYISÉGLOPÁS ÉS IDENTITÁSHAMISÍTÁS ELŐZMÉNYEIRŐL

A tények, információk elferdítése, meghamisítása, majd azok nyilvánosságra hozatala, megtévesztő híresztelése nem újdonság, hanem gyakorlatilag egyidős az emberiséggel. A személyiség meghamisítása, más szerepének felvétele szintén hosszú múltra tekint vissza, már vallási szövegekben is találunk rá példát. Tóth Dávid (2020: 113) a személyiséglopás büntetőjogi elemzéséről szóló tanulmányában utal a bibliai Jákob és Ézsau történetére: Izsák elsőszülött fia, Ézsau lemondott az őt megillető előjogokról az öccse, Jákob javára egy tál lencséért. Erről azonban

az apjuk, Izsák nem tudott. Amikor később Jákob meglátogatta már vak édesapját, hogy megkapja áldását az örökséghez, Ézsau ruháiban és kecskebőrben jelent meg vak apja előtt, hogy Izsák ne vegye észre a csalást (Tóth 2020: 113).

Történelmi adalékot találunk a témához a középkori Angliában, ahol a 15. században uralkodott IV. Edward, akinek a két fia, Edward és Richard nyomtalanul eltűnt. Az uralkodó halála után trónörökös hiányában öccse, III. Richard lett az új király, akit VII. Henrik követett. VII. Henrik uralkodása alatt azonban két trónkövetelő is felbukkant a semmiből. Az egyik, Lambert Simnel, egy közrendű ifjú, a két eltűnt herceg unokaöccsének adta ki magát, Edward Plantagenet néven, miközben az igazi Edward Plantagenet a Towerben raboskodott. A fiatalembert Írországban megkoronázták mint VI. Edward angol királyt, míg VII. Henrik az igazi herceget szabadon bocsájtotta, így bizonyítva a csalást. Simnel ezek után kegyelmet kapott, majd a királyi udvarban mosogatófiúként, később solymászként szolgált (W1; Szmolenszki 2018: 20). A másik trónkövetelő Perkin Warbeck volt, aki a hatalmon lévő király ellenségeivel Írországban sereget gyűjtött, és kétszer is megpróbált partra szállni, hogy megdöntse a király uralmát. A második akciót követően elfogták és börtönbe vetették. Ennek ellenére Perkin továbbra is a királyi udvarban maradhatott, egészen addig, amíg egy szökési kísérlete után felakasztották (Goble 1999; Szmolenszki 2018: 20).

Évszázadokkal később, az információtechnológia előretörésével a 20. század végén és a 21. század elején a személyiséglopás, ezáltal a valóság meghamisítása teljesen új dimenzióba lépett. Ennek számos oka közül az egyik, hogy az emberek saját maguk osztanak meg információkat önmagukról az interneten, többek között fényképeket és videófelvételeket a közösségi oldalakon és más online platformokon. Az interneten ismert emberekről, közszereplőkről is számos kép és videófelvétel kering, akár a napi híreknek és tudósításoknak köszönhetően, amelyek immár digitalizált formában férhetőek hozzá a nagyközönség számára, ezért a vizuális tartalmak feldolgozása (megvágás, retusálás, átdolgozás stb.) számítógépes úton is sokkal egyszerűbb, mint a régebbi technológiával készített felvételeké. Habár a személyiség „ellopására", így a valóság meghamisítására korábban is léteztek különböző technológiák, a valóság eltorzítására, ezáltal egy valós személy cselekedeteinek meghamisítására való képesség a deepfake-technológia néhány évvel ezelőtti megjelenésével óriási „minőségi" ugrást eredményezett a személyiséglopás, egyben a tényhamisítás és álhírterjesztés piacán.

Komoly visszhangot váltott ki a médiában a volt amerikai elnökről, Barack Obamáról készült deepfake-videó, amelyben az elnök (hamisított mása) ironikusan és vulgáris módon hívja fel a figyelmet a technológia veszélyeire (W2). Szintén hamar felkapták a médiában azt a 2022 márciusában készült deepfake-videót, amelyben az ukrán elnök, Volodomir Zelenszkij utasítja az ukrán hadsereget, hogy tegyék le a fegyvert az orosz hadsereg előtt. A videóról hamar kiderült, hogy hamis, és ugyancsak deepfake-technológiával készült (W3).

A deepfake-technológia ugyanis alkalmas arra, hogy olyan audiovizuális tartalmakat hozzanak vele létre, amelyben egy valódi személy olyan dolgokat mond vagy tesz, amit egyébként sosem mondott vagy tett a valóságban (Chesney–Citron 2019: 1753). Az ilyen típusú személyazonossággal való visszaélés alkalmas lehet az emberek és a közvélemény komoly befolyásolására, hiszen a meghamisított felvételen szereplő személy mind az arca, mind a hangja, mind pedig a gesztusai alapján szinte teljesen hasonlít az eredetire. A különbség – néhány nehezen észrevehető vizuális jelen kívül – csupán annyi, hogy az általa előadottak a valóságban az eredeti szereplőtől sosem hangzottak el.

A technológiára építő programok ma már könnyen elérhetők az interneten, alkalmazásuk pedig egyre egyszerűbb, és nem igényel különösebb informatikai tudást. A szórakoztató videótartalmak gyártása mellett a technológia alkalmas arra is, hogy olyan szavakat adjon egy valós személy szájába, illetve olyan tetteket hajtassanak vele végre a képernyőn, amit egyébként nem mondana vagy tenne, ezáltal sértve meg többek között a személyes adatai védelméhez fűződő jogait (Miskolczi–Szathmáry 2018: 141–142). A technológia mind jó, mind rossz célokra alkalmas lehet. A fejezet célja a technológia értékelése adatvédelmi szempontból az Európai Unió általános adatvédelmi rendeletének (közkeletű angol rövidítése szerint: a GDPR) tükrében (W4). Mielőtt azonban rátérnénk a jelenség adatvédelmi értékelésére, nézzük meg először az annak alapjául szolgáló technológia sajátosságait.

2. A DEEPFAKE ALAPJÁUL SZOLGÁLÓ TECHNOLÓGIA: GÉPI TANULÁS, MÉLYTANULÁS

A Jan Kietzmann és szerzőtársai által használt fogalom szerint: „a deepfake a gépi tanulás és a mesterséges intelligencia technológiájának felhasználásával olyan vizuális, illetve hangalapú tartalmakat hoz létre, amelyek nagy eséllyel lesznek megtévesztőek" (Kietzmann et al. 2019: 1).

A deepfake alapja a gépi tanuláson alapuló videó- és hangmanipuláció. A gépi tanulás mint a mesterségesintelligencia-fejlesztés egyik ágának lényege, hogy a rendszer saját maga képes a rendelkezésére álló adatokból, információkból, azaz korábbi tapasztalatokból tudást létrehozni. A rendszer példaadatok, minták alapján képes önállóan vagy emberi segítséggel szabályszerűségeket, szabályokat felismerni és meghatározni, majd az elsajátított tudásbázisban felfedezett szabályszerűségek alapján döntéseket hozni. A gépi tanulás során az emberi programozó csak az önálló tanulás kiindulópontjaként használható adatokat ad meg a számítógépnek. Ezek után a gép maga hozza létre és fejleszti tovább azokat az algoritmusokat, amelyekre a döntések – vagy előrejelzések – meghozatalához szüksége van. Az adatokban felfedezett szabályszerűségekből létrehozza az úgynevezett modellt,

amely tartalmazza az elsajátított mintázatokat, szabályszerűségeket, majd ezeket alkalmazza az új adatok esetén is. A gép a saját maga által felállított modell alapján hoz immár önálló döntést bármilyen emberi beavatkozás nélkül. Az emberi felhasználó már csak a meghozott döntéssel szembesül a program használata során (Datatilsynet 2018: 7–8).

A deepfake-tartalmak létrehozásához gépi tanulási módszereket használnak, annak is egy különleges ágát, amelyet mélytanulásnak *(deep learning)* neveznek. Ez a gépi tanuláshoz képest még szűkebb fogalom. A mélytanulás mesterséges neuronhálózatok alkalmazásán alapszik. Ezeken a mesterséges neuronhálózatokon az emberi neuronok közötti adatátvitel mintájára a számítógépes jelek nem pusztán pozitív és negatív válaszok (tehát 0 és 1) megkülönböztetésére szolgálnak (Hlács 2016).

Az emberi agy működésének alapja a neuronok közötti kapcsolat, azaz a szinapszis. A neuronok a szinapszisokon keresztül küldenek jelet egymásnak, egy neuron pedig párhuzamosan több szinapszist, tehát több kapcsolatot is képes fenntartani másik neuronokkal. A szinapszisok azonban nemcsak a jelek átvitelére képesek, hanem a korábbi kapcsolatok emlékét is meg tudják őrizni. Az agyban gyakrabban használt szinapszisokban a kapcsolatok erősebbek lesznek, ezért van az, hogy egyes gondolatok könnyebben fogalmazódnak meg és jutnak eszünkbe, vagy bizonyos problémát gyorsabban, hatékonyabban tudunk megoldani. A mélytanulás is hasonlóan működik: a többrétegű adatbázisban a számítógép feldolgozó egységei egyrészt végzik a 0 és 1 válaszok megkülönböztetését, továbbá a többi réteg ugyanazon kérdésére adott válaszát is vizsgálják. Így az adott kérdésre minden egyes rétegben létrejön a 0 vagy az 1 válasz, azonban a rétegeket összevetve az adott válasz tényleges értéke súlyozható (Alpaydin 2016: 85–86; Schiff 2020: 16). Az algoritmus itt is az általa beazonosított mintázatok alapján próbálja meg előre jelezni a következő lehetséges értékeket, viszont a mélytanulást végző neuronhálózat többrétegű (általában már tanított) adatbázisból és rájuk épülő tanuló algoritmusokból épül fel. Ennek köszönhetően a rendszer absztrakciós készsége is növekszik a megoldásra váró feladatokkal kapcsolatban, így ez a technológiai megoldás sokkal összetettebb és gyorsabban fejlődő algoritmusokat, pontosabb válaszokat és magasabb fokú absztrakciót tesz lehetővé az MI számára (Hlács 2016).

A deepfake-videós tartalmak készítése általában generatív neurális hálózati architektúrák, legtöbbször úgynevezett generatív ellenséges hálózatok (GAN-ok) betanítását foglalja magában (Chesney–Citron 2019: 1756–1757). A GAN-módszert a Google egyik kutatója, Ian Goodfellow és szerzőtársai dolgozták ki 2014-ben (Goodfellow et al. 2014). A GAN két neuronhálózatot kapcsol össze, hogy azok párhuzamosan dolgozzanak. Az egyik hálózat, a „generátor" a rendelkezésre álló adathalmazból kiválogatja azokat a már meglévő, kész képeket, amelyek a célvideón szereplő arc tulajdonságaira (szög, arckifejezés, árnyéko-

lás stb.) a leginkább hasonlítanak. Ezek után a másik hálózat, az úgynevezett „diszkriminátor" elemzi a két adathalmaz hasonlóságait és eltéréseit. Az esetleges eltéréseket a számítógép automatikusan korrigálja, hogy a kép minél jobban hasonlítson az eredeti videón szereplő arc tulajdonságaira (módosítja az arc színeit, mimikáját, a száj, a szemek és az arcizmok mozgását stb.). A diszkriminátor hálózat folyamatosan kommunikál a generátorral, és szükség esetén újabb képeket használ fel a meggyőzőbb eredmény elérésére (Chesney–Citron 2019: 1760–1762; Schiff 2020: 18–19).

A GAN tehát egy valódi többrétegű mélytanuló neuronhálózatra épülő rendszer, amely önmagát is folyamatosan fejleszti, ahogy egyre több adatot (képet) elemez és hoz létre. A GAN képes a saját maga által korábban készített képekből is tanulni, így egyre szofisztikáltabb eredményeket produkál.

3. A DEEPFAKE-TECHNOLÓGIA ALKALMAZÁSÁNAK KAPCSOLATA A SZEMÉLYES ADATOKKAL

A deepfake-technológia kapcsolatát a személyes adatokkal két külön szakaszra érdemes bontani: az első a technológiát alkalmazó szoftver tanítása, a második pedig annak alkalmazása.

A tanítás során a rendszerbe először nagy mennyiségű adatot táplálnak be, amelyben a tanuló algoritmus megpróbál mintákat, hasonlóságokat keresni. Amennyiben talál ilyen azonosítható mintákat, úgy azokat megjegyzi, és elmenti későbbi használat céljából. A megjegyzett és elmentett minták alapján alkotja meg a rendszer az ún. modellt (Datatilsynet 2018: 7). Amennyiben a tanításra használt adathalmaz személyes adatokból épül fel (ilyenek lehetnek az emberek arcát, mozgását vagy hangját tartalmazó képek és hangok), úgy a tanítás szükségszerűen együtt jár személyes adatok kezelésével is.

A már tanított szoftver alkalmazása mint második lépcső szintén együtt járhat személyes adatok kezelésével, hiszen amennyiben a rendszerbe újabb adatokat töltenek fel, amelyek hasonlóak a tanuláshoz használt adatokhoz, a modell alapján az eldönti, hogy az új adat mely megtanult mintázathoz hasonlít a leginkább, s később ez alapján fogja az elsajátított mintázatok alapján kialakítani az automatikus döntést (Datatilsynet 2018: 7).

A deepfake-tartalmak előállítása során egy, a program által létrehozott „szereplő" rendelkezik olyan tulajdonságokkal, amelyek egy létező természetes személytől vagy személyektől származnak. A deepfake-videó szereplőjének az arcvonásai, a hangja, de akár a mimikája is az emberi szem számára felismerhetetlenül hasonlíthat arra a természetes személyre, akiről a deepfake-videó szereplőjét mintázták. Ezen tartalmak elkészítéséhez és a használt szoftverek működéséhez kimondottan sok személyes adatra van szükség. A deepfake-tartalom minősége annál jobb, te-

hát az azon látható szereplő annál életszerűbb, minél több és minél jobb minőségű adat alapján készült (Klein–Tóth 2018: 67–68).

A tanításhoz használt személyes adatok, a szoftver által a deepfake-tartalom előállításához használt személyes adatok és végül a kész tartalom mint személyes adat jogi értékelése adatvédelmi szempontból elkülönül egymástól. A tanítás, a tartalom előállítása és végül maga a tartalom közzététele külön-külön adatkezelési műveletek, és jogszerűségüket ennek megfelelően érdemes külön vizsgálni (Schiff 2020: 25). A továbbiakban ezen három lépés alapján mutatom be az egyes adatkezelési műveletek adatvédelmi kérdéseit.

4. A DEEPFAKE-TARTALMAKAT ELŐÁLLÍTÓ SZOFTVER TANÍTÁSÁHOZ HASZNÁLT SZEMÉLYES ADATOK

A szoftver tanításának adatvédelmi vizsgálata szempontjából szükséges felvázolni néhány alapfogalmat. Az első kérdés az, hogy melyek azok a személyes adatok, amelyek a deepfake-tartalmak előállítására használt szoftver tanításakor jelentőséggel bírhatnak.

Ha a GDPR vonatkozó fogalmait keressük, akkor az első mindenképpen a személyes adat fogalma. Eszerint személyes adatnak minősül az azonosított vagy azonosítható természetes személyre – a GDPR szóhasználatával élve: az érintettre – vonatkozó bármely olyan információ, amely alapján ő akár közvetlen, akár közvetett módon azonosítható (GDPR 4. cikk 1. pont). A joggyakorlat is többször foglalkozott már az érintettről készült kép- és hangfelvétel személyesadat-jellegével. Ez alapján határozottan elmondható, hogy egy ember arca, képmása személyes adatnak, a képfelvétel készítése, valamint az adatokon elvégzett bármely művelet pedig adatkezelésnek minősül (NAIH 2022: 10). Így a természetes személyekről készült kép- és hangfelvételeknek a szoftverek általi feldolgozása (például az algoritmus tanításának céljára) adatkezelést fog eredményezni.

A GDPR hatálya alá tartozó adatkezelések kapcsán fontos kitétel, hogy a személyes adatok kezelésének részben vagy egészben automatizált módon kell történnie, ahhoz, hogy arra alkalmazni kelljen a rendelet előírásait. A GDPR-t ezenfelül azoknak a személyes adatoknak a nem automatizált módon történő kezelésére is alkalmazni kell, amelyek valamely nyilvántartási rendszer részét képezik, vagy amelyeket egy nyilvántartási rendszer részévé kívánnak tenni, bár ezen utóbbi szabály témánk szempontjából kevésbé bír jelentőséggel [GDPR 2. cikk (1) bekezdés].

Mind a gépi tanulás, mind az ennek szűkebb szeletét jelentő és a deepfake-technológia alapját adó mélytanulás során létrejövő adatkezeléshez jellemzően sok személyes adatra van szükség. A szoftver megtanítása a feldolgozott személyes adatok, így a kép-, videó- és hangfájlok révén arra, hogy valódinak tűnő

tartalmakat tudjon előállítani, jellemzően sok bemeneti tanító adatot igényel. A felhasználandó adatok mennyiségével a legtöbb esetben egyenesen arányos a létrehozott deepfake-videó minősége is, hozzátéve azt is, hogy a legmegfelelőbb adatok előzetes kiválasztása és címkézése is ugyanilyen fontos szempont a hatékony tanítás elérésére (Datatilsynet 2018: 10).

A deepfake-tartalmak előállítására alkalmas szoftver tanításához felhasznált személyes adatokért való adatkezelői felelősség szintén fontos problémakör. A GDPR fogalomrendszerében adatkezelőnek minősül az a természetes vagy jogi személy, közhatalmi szerv, ügynökség vagy bármely egyéb szerv, amely a személyes adatok kezelésének céljait és eszközeit önállóan vagy másokkal együtt meghatározza (GDPR 4. cikk 7. pont). A deepfake-videók előállítására alkalmas szoftver tanítása szempontjából ezért adatkezelőnek azon személyt (akár természetes, akár jogi személyről beszélünk) vagy szervezetet kell tekinteni, aki vagy amely a tanítás alapjául szolgáló adatbázis összeállításának céljait és eszközeit, továbbá ezen adatbázis felhasználásának céljait és eszközeit előzetesen meghatározta. A gyakorlatban ez legtöbbször egy szoftverfejlesztéssel foglalkozó vállalkozást takar, amely abból a célból állít össze személyes adatokból egy tanító adatbázist, hogy azt mélytanuló algoritmusok segítségével elemeztesse egy szoftverrel avégett, hogy a beazonosított mintázatok alapján az később minél élethűbb deepfake-videókat tudjon előállítani.

Az adatkezelői minőség tehát az adatkezelés tanítási stádiumában megáll a személyes adatok tanításra való felhasználásának műveleténél. A szoftverfejlesztő adatkezelői felelőssége a tanító adatbázis összeállításánál, a tanítás céljainak meghatározásánál és az ezen célok elérése érdekében megírt tanuló algoritmusok kiválasztásánál vagy megírásánál ér véget. A tanítás és a szoftverfejlesztés során ezért a fejlesztővállalkozásnak meg kell felelnie a GDPR előírásainak. Ehhez képest magának a konkrét deepfake-videónak az elkészítése és ehhez egy megszemélyesíteni kívánt személy személyes adatainak összegyűjtése és felhasználása már elkülönült adatkezelői minőséget fog fő szabály szerint eredményezni.

A teljesség igénye nélkül az alábbiakat szükséges figyelembe venni a tanításkor. A gépi tanuláson alapuló MI fejlesztése során is figyelembe kell venni az olyan, a GDPR-ban is meghatározott alapelveket, mint az adatok kezelésének célhoz kötöttsége, szükségesség-arányossága és a megfelelő adatkezelési jogalap megléte (GDPR 5. cikk).

Különös figyelemmel kell lenni továbbá ezen programoknál a beépített, illetve az alapértelmezett adatvédelem elvére is (GDPR 25. cikk). Az adatok feldolgozása csak átláthatóan és elszámoltatható módon történhet, valamint az érintetti jogok gyakorlása sem korlátozható. Az adattakarékosság elvének érvényesülésére kifejezetten javasolt szintetikus – valós adatokból mesterségesen generált (konkrét személyhez nem köthető) – adatok felhasználása (Datatilsynet 2018: 25). A szintetikus adatok használatával elkerülhető az, hogy a szoftverfejlesztő esetleg megsértse

a GDPR vonatkozó előírásait, mivel a konkrét személyhez nem köthető, mesterségesen generált adatokra már nem kell alkalmazni a felhasználásuk során a rendelet előírásait, lévén annak a hatálya csupán a konkrét, élő természetes személyhez köthető adatokra terjed ki (Péterfalvi–Révész–Buzás 2021: 51, 71).

5. A DEEPFAKE-TARTALOM ELKÉSZÍTÉSÉHEZ HASZNÁLT SZEMÉLYES ADATOK

A konkrét deepfake-videók elkészítéséhez a szoftver általános tanításán túl szükség van minden esetben olyan bemeneti adatokra, amelyek elemzése révén a szoftver elő tudja állítani magát a hamisított tartalmat. Egy ismert közéleti szereplőről, például egy ország miniszterelnökéről készítendő deepfake-videóhoz a szoftvernek össze kell gyűjtenie olyan fotókat, videókat és hangokat, amelyek valóban az adott alanyról készültek. Ezek elemzése révén a már a tanítás során beazonosított általános mintázatok és a konkrét személyre vonatkozó valódi tartalmak alapján lehet elkészíteni a hamisított felvételt.

A konkrét tartalom tehát úgy készül el, hogy a szoftver a tanítása során megtanulja azt, hogy általánosságban miként néznek ki az emberek, és egyes szavak formálása során milyen általános mimikai jelek figyelhetőek meg az arcon. Ezek után a szoftver használata során kiválaszt a felhasználó egy személyt, akiről hamisított felvételt szeretne elkészíteni, ehhez pedig a konkrét személyről készített valódi felvételeket tölt fel a programba. A konkrét személy arcának, hangjának, mimikájának és más testi tulajdonságainak elemzése és az általános mintázatok alapján végül elkészül a konkrét deepfake-videó.

Természetesen ehelyütt is érdemes elmondani, hogy a deepfake-videó elkészítéséhez a megszemélyesíteni kívánt érintettről felhasznált tartalmak, így képek, videók, hangok is személyes adatnak fognak minősülni a GDPR vonatkozó fogalmai alapján, mivel – ahogy az az előző pontban is kifejtettem – egy ember képmása személyes adatnak, az azokon elvégzett bármely művelet pedig adatkezelésnek minősül (GDPR 4. cikk 1. és 2. pontok). A deepfake-videó készítésekor ezért a megszemélyesíteni kívánt valódi személy adatainak kezelése során is meg kell felelni a GDPR előírásainak fő szabály szerint. A személyes adatok ilyen célú felhasználására ezért elvileg legitim céllal és megfelelő jogalappal kell rendelkeznie az azt előállító tartalomkészítőnek.

Ez a gyakorlatban azt jelenti, hogy egy adott személyről egy deepfake-videó készítéséhez a videót készítő személynek tudnia kell azt a GDPR 5. cikk (2) bekezdésében foglaltak, azaz az úgynevezett elszámoltathatóság elve alapján igazolni, hogy megfelelő jogalappal rendelkezik a személyes adatok ilyen célú kezeléséhez, továbbá a GDPR egyéb előírásait is betartja az adatkezelés során. Például az ilyen videók készítőjének be kell tudnia azt mutatni a GDPR 7. cikk (1) bekezdése alap-

ján, hogy beszerezte az érintettől a GDPR 6. cikk (1) bekezdés a) pontja szerinti hozzájárulását egy ilyen célú adatkezeléshez. Tehát az érintettnek hozzá kellene ahhoz járulnia, hogy a személyes adatai felhasználásával deepfake-videót fognak róla készíteni. Ezzel kapcsolatban a videó készítőjének pedig előzetesen és megfelelő tartalommal tájékoztatnia kell őt az adatkezelés körülményeiről a GDPR 13. cikkének megfelelően.

Lássuk be, az előzőekben kifejtett adatvédelmi előírásoknak való megfelelés a deepfake-videók készítése során az esetek túlnyomó többségében egyáltalán nem tud érvényesülni, hiszen a hamis videóban megszemélyesített személyről legtöbbször lejáratási, megtévesztési céllal készül a felvétel, annak elkészítéséhez pedig nyilvánosan elérhető tartalmakat használnak fel. Ezért nemcsak maga a már elkészült hamis felvétel nyilvánosságra hozatala, hanem annak készítéséhez az eredeti személyes adatok felhasználása is túlnyomórészt jogsértő lesz. Természetesen el lehet képzelni olyan esetet is, amikor valakinek a tudomásával és beleegyezésével, így legitim adatkezelési céllal és jogalappal készül a személyes adatai felhasználásával egy deepfake-videó (például oktatási vagy tudatosítási célból), azonban ez valószínűsíthetően az esetek elenyésző hányadát teszi ki.

6. A DEEPFAKE-VIDEÓ MINT SZEMÉLYES ADAT

A már elkészült deepfake-videó személyesadat-jellegével kapcsolatban a legfőbb értelmezési kérdés, hogy tulajdonképpen az abban szereplő „személy" soha nem tanúsította a videón látható magatartást, illetve nem hagyták el a száját az elhangzó szavak, mondatok. Ez esetben tehát egy olyan mesterségesen generált tartalom személyesadat-jellegét kellene megállapítani, amely nem az érintettről készült, nem is tőle származik, hiszen magán a videón nem is ő szerepel, hanem egy szintetikus entitás. Kérdés tehát, hogy olyan elkülönült jogi sorsot osztó tartalomnak kell-e tekinteni a deepfake-videót, amely nem minősül az utánzott érintett személyes adatának, vagy inkább a testi és más külső jellemzők nagyfokú hasonlósága miatt a tartalom az utánozni kívánt személy személyes adata lesz-e (Schiff 2020: 28).

A GDPR meghatározása alapján a személyes adat fogalmának kulcseleme az, hogy az alapján egy természetes személy azonosítható legyen, akár közvetett, akár közvetlen módon. Azonosíthatóvá pedig egy természetes személy a vele kapcsolatba hozható bármely információ alapján válhat, például ilyen információk lehetnek a testi, fiziológiai vagy szociális azonosságára vonatkozó tényezők (GDPR 4. cikk 1. pont). Ezen fogalom alapján elmondható, hogy nem feltétlenül szükséges az egy természetes személy azonosíthatóságának megállapításához, hogy az információ közvetlenül tőle származzon. Elég az, ha egy konkrét, tehát élő természetes személyre rá lehet ismerni a közölt információk alapján, amelyek testi, fiziológiai

jellemzők is lehetnek. Például egy ismert személy jellegzetes hanghordozására, testtartására, arcának és mimikájának jellemzőire vonatkozó információ feldolgozása révén előállított szintetikus deepfake-tartalom a GDPR alapján a megszemélyesíteni kívánt érintett személyes adatának fog minősülni. Ilyen módon tehát még egy olyan tulajdonság vagy jellemző is minősülhet személyes adatnak, amely a szemlélőkben azt az érzést kelti, hogy ők ezt a tulajdonságot egy meghatározott élő természetes személyhez tudják kötni. Az információk tehát nemcsak objektívek, hanem akár szubjektívek is lehetnek, és nem szükséges az sem, hogy valósak legyenek. Tehát személyes adatnak minősülhet akár valótlan vagy nem igazolt információ is. Ezért a deepfake-videó is a megszemélyesített, így mások által már beazonosítható, tehát egy adott csoport (vagy az egész társadalom) minden más tagjától elkülöníthető természetes személy személyes adata lesz, ezért alkalmazandók rá a GDPR előírásai (Balogh et al. 2019: 48–49). Ez az értelmezés már csak azért is tűnik logikusnak, mivel bizonyos deepfake-videók készítése és nyilvánosságra hozatala az azokon keresztül megvalósuló „személyiséglopás" miatt képes súlyosan negatívan befolyásolni az érintett magánszféráját, ezen belül jó hírnevét, társadalmi megítélését, magán- és családi életét, sőt akár hátrányos megkülönböztetés oka vagy indoka is lehet vele szemben, beleértve akár a lehetséges vagyoni és nem vagyoni károkat is. Az ilyen tartalmak nyilvánosságra hozatala a magánszférára gyakorolt hatásuk miatt olyan alapvető kockázatokat hordoz az egyénre nézve, amelyeket maga a GDPR is külön nevesít a 75-ös preambulumbekezdésében.

Az előző ponthoz hasonlóan tehát a konkrét deepfake-videó kapcsán is elmondható, hogy annak kezelésére, így például a nyilvánosságra hozatalára csak akkor kerülhetne sor, ha ehhez jogszerű céllal és jogalappal rendelkezik az azt nyilvánosságra hozó személy. Tehát az ilyen videók nyilvánosságra hozójának is elvileg be kellene tudnia azt mutatni, hogy beszerezte az érintettől annak hozzájárulását egy ilyen célú adatkezeléshez, és őt előzetesen az adatkezelésről tájékoztatta is. Sajnos azonban itt is elmondható, hogy az esetek túlnyomó többségében GDPR ezen előírásai szintén nem tudnak érvényesülni, hiszen a megszemélyesített személyről legtöbbször lejáratási, megtévesztési céllal hozzák nyilvánosságra az elkészült hamis felvételt. Természetesen ilyen esetekben is elképzelhető legitim adatkezelés, szintén oktatási célokból, de az esetek túlnyomó részét nem az ilyen tartalmak teszik ki.

7. ADATVÉDELMI JOGÉRVÉNYESÍTÉSI LEHETŐSÉGEK A DEEPFAKE-TARTALMAK KÉSZÍTÉSE ÉS NYILVÁNOSSÁGRA HOZATALA KAPCSÁN

A deepfake-videó készítése, illetve nyilvánosságra hozatala jogellenesen történik, ha nem rendelkezik az ezt véghez vivő adatkezelő a GDPR szerinti megfelelő, a 6. cikk (1) bekezdése alapján is igazolható jogalappal. Ilyen jogalap lehet az érintett hozzájárulása a személyes adatainak egy vagy több konkrét célból történő kezeléséhez, szerződéskötés teljesítése, az adatkezelőre vonatkozó jogi kötelezettség teljesítése, természetes személy létfontosságú érdekeinek védelme közérdek vagy közhatalmi jogosítvány gyakorlása vagy a jogos érdek.

A GDPR alkalmazásában a megfelelő jogalap megléte önmagában még nem elegendő ahhoz, hogy az adatkezelés jogszerűnek legyen tekinthető, ugyanis bármely olyan tényező, illetve körülmény, amely miatt a személyes adatok kezelése nem felel meg az adatvédelmi rendeletnek, jogellenességet eredményez (Péterfalvi–Révész–Buzás 2021: 181–183). Így például hiába adta az érintett kifejezett hozzájárulását a személyes adatai kezeléséhez, ha egyébként az adatkezelést megelőzően részére nyújtott tájékoztatás nem volt teljes körű, és ezért nem felel meg a GDPR érintetti tájékoztatással kapcsolatos, a 13. cikkben található előírásainak. A rendelet tehát ebben az esetben is jogellenesnek fogja tekinteni az adatkezelést, hiszen abból a logikából indul ki, hogy megfelelő tájékozottság hiányában nem jöhet létre valóban önkéntes hozzájárulás. Amennyiben az érintett tehát nem tudja, hogy pontosan mihez járul hozzá, úgy nem beszélhetünk a hozzájárulására alapított jogszerű adatkezelésről sem.

A személyes adatok kezelésének a jogellenességét az EU-ban az adott tagállam adatvédelmi felügyeleti hatósága és bíróság is megállapíthatja a GDPR 77. és 79. cikkei alapján. A felügyeleti hatóság legtöbbször a jogellenes adatkezelésről az érintett panasza vagy egy a jogsértést észlelő harmadik személy bejelentése alapján értesül, de előfordulhat, hogy a hatóság a sajtóból vagy az internetről értesüljön a visszaélésről. A bíróság szintén az érintett kereseti kérelme alapján szerezhet tudomást a jogsértésről, vagy ha az adatvédelmi felügyeleti hatóság a jogsértést megállapító közigazgatási határozatát a GDPR 78. cikkében foglaltak alapján megtámadják a bíróság előtt.

A probléma a GDPR által biztosított fenti jogorvoslati rendszerrel, hogy a deepfake-videókat készítő és közzétevő adatkezelők beazonosítása a legtöbb esetben rendkívül nehéz feladat, legalábbis az adatvédelmi hatóságok közigazgatási eljárásainak keretei között. Ennek oka, hogy a hamis videók készítői általában ügyelnek arra, hogy nehezen legyenek beazonosíthatók az interneten. A tényleges jogsértést elkövető személy mint adatkezelő beazonosítása ezért a büntetőeljárásokban használatos nyomozati eszközöket igényelne.

A hatósági vagy bírósági úton történő jogérvényesítés eszközein túl az érintettnek is lehetősége van, hogy akár közvetlenül az adatkezelőhöz forduljon, és a személyes adatai jogellenesen kezelése esetén kérje azok törlését a tartalom készítőjétől vagy közzétevőjétől a GDPR 17. cikk (1) bekezdés d) pontja alapján. Természetesen ezen érintetti joggyakorlás kapcsán is el lehet mondani, hogy a legtöbb esetben nem hatékony módszer a deepfake-tartalmakkal összefüggő jogsértések orvoslására, mivel az adatkezelőnek nem áll érdekében azok teljesítése, sőt legtöbbször nem is reagálnak ezekre (ha egyáltalán elérhetőek valamilyen címen keresztül). A deepfake kapcsán felmerülő jogsértések kivizsgálása ezért az érintettek rendelkezésére álló, továbbá hatósági és bírósági eszközökkel a legtöbb esetben sajnos egyáltalán nem hatékony.

További megoldás lehet a deepfake készítője, illetve közzétevője ellen büntetőeljárás indítása, mivel a legtöbb esetben az ilyen tartalmak a Büntető Törvénykönyvről szóló 2012. évi C. törvényben foglalt, a 226/A. §-ba ütköző „a becsület csorbítására alkalmas hamis hang- vagy képfelvétel készítésének" vagy a 226/B. § szerinti ilyen felvétel „nyilvánosságra hozatalának" tényállását meríthetik ki. E bűncselekmények speciális elkövetési magatartása a hamis, hamisított vagy valótlan tartalmú hang- vagy képfelvétel készítése, nyilvánosságra hozatala vagy hozzáférhetővé tétele, így azok tárgyába a deepfake-videók is beletartozhatnak. Jelen fejezetnek nem képezi szűk értelemben vett tárgyát a deepfake jelenségnek az adatvédelmi mellett a büntetőjogi értékelése is, ezért ennek további elemzésétől a tartalmi keretek miatt eltekintek.

8. JOGALKOTÁSI IRÁNYOK A DEEPFAKE-TARTALMAK ÁLTAL A MAGÁNSZFÉRÁRA JELENTETT KOCKÁZATOK KEZELÉSE ÉRDEKÉBEN

A már elkövetett, deepfake-tartalmakkal kapcsolatos konkrét jogsértések elleni érintetti és hatósági fellépés nehézségeit az elmúlt időszakban a jogalkotás is felismerte, és igyekszik hatékonyabb eszközöket biztosítani. Ebben a pontban a deepfake-tartalmak elleni küzdelemmel kapcsolatos legújabb európai uniós jogalkotási javaslatokat tekintem át.

Az első jogalkotási javaslat, amely foglalkozik a deepfake-technológia szabályozásával, az Európai Bizottság által nyilvánosságra hozott, a mesterséges intelligenciát érintő, ún. MI-rendelettervezet (W5), amely valamennyi uniós tagállamban egységesen végrehajtandó jogszabályként szabályozná a mesterséges intelligencia fejlesztését. A tervezet az MI-ként való besorolásra három feltétel együttes teljesülését írja elő. Először is az MI-nek meghatározott technológiákat kell alkalmaznia, másodszor az ember által kijelölt célokat önállóan kell tudnia követni, végül olyan kimeneteleket kell tudnia produkálni, amelyekkel „befolyásolja" a környezetet. Az

új kódex tervezete a gépi tanuláson alapuló rendszerekre is kiterjeszti a hatályát (MI-rendelettervezet 3. cikk 1. pont és I. melléklet).

A rendelettervezet kockázatalapú megközelítést alkalmaz az MI-k besorolása szempontjából, amely összesen négy nagy kategóriába igyekszik felosztani a rendszereket. Ezek alapján különbséget tesz az elfogadhatatlanul magas kockázatúként besorolt, azaz tiltott rendszerek, a magas kockázatú rendszerek, a korlátozott kockázatú rendszerek és a minimális/kockázat nélküli rendszerek között.

Anélkül, hogy mélyebben elemeznénk az egyes kockázati szinteket, témánk, tehát a deepfake-technológia adatvédelmi értékelése kapcsán fontos azt megemlíteni, hogy az MI-rendelettervezet kifejezetten utal a szövegében többször is erre a technológiára. A javaslat indokolása kiemeli, hogy célja megbízható kockázati módszertant meghatározni az olyan MI-rendszerek értékelésére, amelyek jelentős kockázatot jelentenek az emberek egészségére és biztonságára vagy alapvető jogaira nézve. Ezeknek az MI-rendszereknek meg kell felelniük a megbízható mesterséges intelligenciára vonatkozó kötelező horizontális követelményeknek, és megfelelőségértékelési eljárások tárgyát kell képezniük, mielőtt forgalomba hoznák őket az uniós piacon. Ezek célja, hogy az MI-rendszerek teljes életciklusa során biztosítsák a biztonságot és az alapvető jogok védelmét biztosító meglévő jogszabályok tiszteletben tartását. A javaslat egyes MI-rendszerek esetében azonban csak minimális átláthatósági kötelezettségeket javasol. Idetartoznak különösen a csevegőrobotok vagy a deepfake-tartalmak (MI-rendelettervezet indokolása 1.1. pont).

A deepfake-tartalmakkal kapcsolatban a rendelettervezet megemlíti, hogy szabályozásukra az általuk jelentett manipulációs kockázatok miatt van szükség, mivel azok erre alkalmas tartalmakat hoznak létre, vagy meglévő tartalmakat manipulálnak. Ezért a tervezet szerint, ha az emberek ilyen rendszerekkel érintkeznek, úgy őket előzetesen tájékoztatni kell annak tényéről, hogy deepfake-kel van dolguk. Ezt a rendelettervezet úgy fogalmazza meg, hogy ha az MI-rendszert olyan képek, audió- vagy videótartalmak létrehozására vagy manipulálására használják, amelyek érzékelhetően hasonlítanak hiteles tartalmakra, kötelezővé kell tenni annak előzetes közlését, hogy a tartalom mesterséges módon jött létre, vagy azt manipulálták [MI-rendelet-tervezet 52. cikk (3) bekezdés]. Ez lehetővé teszi a személyek számára, hogy megalapozott döntéseket hozzanak, vagy visszalépjenek adott helyzetből.

A rendelettervezet tehát kötelezővé tenné a deepfake-tartalmak kapcsán annak előzetes közlését, hogy az ilyen technológiát használva, tehát a „valóság manipulálásával" jött létre. Ennek hiányában a deepfake-tartalom valószínűsíthetően jogellenes lesz, így annak létrehozása és nyilvánosságra hozatala nemcsak adatvédelmi, hanem MI-szabályozási szempontból sem lesz jogszerű.

A szabályozás előremutató, azonban a legfőbb kérdés annak hatékonyságával kapcsolatban, hogy vajon miként lesz majd képes azt az EU betartatni. Erre a vá-

laszt szabályozási szempontból azonban nem feltétlenül az MI-rendelettervezetben, hanem egy másik, még alkalmazás előtt álló jogszabályban, a Digitális Szolgáltatások Egységes Piacáról szóló rendelettervezetben *(Digital Services Act, DSA)* találjuk (W6).

A DSA kidolgozására azért volt szükség, mivel az online szolgáltatások biztosításával kapcsolatos jelenlegi szabályozás a 2000-ben hatályba lépett, az elektronikus kereskedelemről szóló irányelvben található, amelynek elfogadására immár több mint két évtizede került sor (W7). Az online világ azóta hatalmasat változott, az online platformokon keresztül történő kommunikáció és a határokon átnyúló online kereskedelem a mindennapok részévé vált. A DSA hatálya éppen ezért a legtöbb online elérhető szolgáltatásokat kínáló platformra kiterjed, beleértve az online piactereket, kereskedelmi platformokat, közösségi oldalakat, tartalommegosztó szolgáltatásokat, applikációk letöltését kínáló weblapokat. Ennek oka, hogy ezeket a felületeket gyakran használják illegális tartalmak terjesztésére vagy illegális áruk vagy szolgáltatások online értékesítésére (W8). Néhány nagy szereplő pedig olyan meghatározó – és ez idáig nem vagy nem kielégítően szabályozott – befolyásra tett szert az online piacok és az online információcsere területén, ami komoly nemzetbiztonsági és alkotmányos aggályokat vet fel (lásd a Cambridge Analytica-botrányt; Domokos 2021: 122).

Erre kíván megoldást nyújtani a DSA többek között a jogellenes online tartalmak eltávolítására vonatkozó kötelező szabályok bevezetésével, a szolgáltatók online piacokon történő nyomon követhetőségével, továbbá a tartalommoderáció egységes és általános szabályozásának előírásával és átláthatósági intézkedésekkel (algoritmusok, ajánlórendszerek működése, célzott hirdetések szempontjai, a hirdetők azonosíthatósága). Ezek kapcsán a platformoknak többek között olyan intézkedéseket kell bevezetniük, amelyek kapcsán ki tudják szűrni a manipulatív és hamisított tartalmakat, beleértve a megtévesztő deepfake-videókat is (Moyer 2022). Ennek gyakorlati kivitelezése természetesen további kérdéseket vet fel, egyes álláspontok szerint – a tartalmak automatikus elemzése és szűrése mellett – az emberi beavatkozást is lehetővé tevő moderálási rendszerek bevezetésére és a panasztétel jogának biztosítására is szükség lesz a DSA vonatkozó rendelkezéseinek való megfelelés érdekében (Frosio–Geiger 2022: 37–38). A DSA-t 2024. január 1-től kellene alkalmazni az Európai Unióban valamennyi, a hatálya alá tartozó szolgáltatónak.

9. ÖSSZEGZÉS

Mint az előzőekben láthattuk, a deepfake-tartalmak előállítására alkalmas MI-szoftverek tanítása és az ilyen tartalmak előállítása személyes adatok kezelésével jár, azonban ezen adatkezelések jogszerűsége a legtöbb esetben erősen kérdéses. Ezenkívül magának a már kész tartalomnak, tehát a deepfake-videónak mint sze-

mélyes adatnak a felhasználása is számos jogsértő helyzetet eredményezhet a mindennapokban.

A jelenleg hatályos vonatkozó uniós jogszabály, tehát a GDPR szabályai kapcsán felmerülő adatvédelmi aggályokon kívül érdemes azonban egy lépéssel hátrébb lépni és általános társadalmi problémaként tekinteni a deepfake-tartalmakra. Az ilyen videók nagy része eleve manipulációs vagy lejáratási céllal készül. A megszemélyesített, szintetikus módon előállított, egyébként egy valós személyre megszólalásig hasonlító szereplő és általa mondott szintén mesterségesen előállított mondatok és tanúsított cselekmények célja a nézők becsapása, manipulálása, a személyiség ellopása és ezen keresztül a valóság meghamisítása. Az ilyen valótlan, az embereket félrevezető tartalmaknak komoly hatásuk lehet az egy-egy társadalmi jelenség kapcsán kialakult közbeszédre és a jelenségre adott társadalmi reakciókra. A meghamisított információk alapján kialakuló közbeszéd és vélemények rossz irányba terelhetik a témáról folyó diskurzust, ami szintén fals, az adott problémát rosszul kezelő reakciókat és megoldásokat szül.

Mindezek miatt nagyon fontos társadalmi kérdés a deepfake-tartalmak megfelelő kezelése és a hatékony jogi reakciók kialakítása azokkal kapcsolatban. Láthatjuk, hogy a jelenleg hatályos és alkalmazandó európai uniós adatvédelmi keretrendszer csak tüneti úton képes kezelni a deepfake által jelentett jogsértéseket, és azokra nem tud igazán hatékony reakciót adni, mivel a tartalmat előállító és terjesztő adatkezelő személyének azonosítása nehezen lehetséges az interneten. A kiváltó okokat az adatvédelmi szabályozás így jelenlegi formájában nem képes hatékonyan kezelni. Örvendetes viszont, hogy a legújabb európai uniós jogalkotási kezdeményezések, így az MI-rendelet és a DSA már konkrétabb fellépést fognak lehetővé tenni. A kérdés már csak az, hogy a deepfake-tartalmak technikai felderítése és azonosítása hatékonyan megvalósítható-e, és így a szolgáltatók érvényt is tudnak-e megfelelően szerezni a jogi előírásoknak. Remélhetőleg, ha talán nem is teljes körű, de mindenképpen magasabb szintű védelem fog létrejönni az elkövetkező időszakban ezek ellen az alapvetően káros tartalmak ellen.

SZAKIRODALOM

Balogh Gyöngyi – Bíró János – Deák Ferenc – Kovács Melinda – Tömösi Ramóna 2019: *Az adatvédelmi jog alapelvei, fogalmai, szereplői, profilalkotás, a személyes adatok különleges kategóriái, bűnügyi személyes adatok*. Budapest: Nemzeti Közszolgálati Egyetem.

Chesney, Bobby – Citron, Danielle 2019: Deep Fakes: A Looming Challenge for Privacy, Democracy and National Security. *California Law Review*, 107/6. https://doi.org/10.15779/Z38RV0D15J [2022. 09. 30.]

Domokos Márton 2021: Globális törésvonalak – a Cambridge Analytica-ügy. In: Szabó Endre Győző (szerk.): *Az Infotörvénytől a GDPR-ig*. Budapest: Ludovika Egyetemi Kiadó. 119–142.

Ethem, Alpaydin 2016: *Machine learning: the new AI*. Cambridge MA: MIT Press Essential Knowledge Series.

Frosio, Giancarlo – Geiger, Christophe 2022: Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime. *European Law Journal* (forthcoming). http://dx.doi.org/10.2139/ssrn.3747756 [2023. 02. 14.]

Goodfellow, Ian J. – Pouget-Abedie, Jean – Mirza, Mehdi – Xu, Bing – Warde-Farley, David – Ozair, Sherjil – Courville, Aaron – Bengio, Yoshua 2014: Generative Adversarial Networks. *Département d'informatique et de recherche opérationnelle, Université de Montréal*. arXiv:1406.2661 [2023. 02. 14.]

Kietzmann, Jan – Lee, Linda W. – McCarthy, Ian P. – Kietzmann, Tim C. 2020: Deepfakes: Trick or treat? *Business Horizons*, 63/2. doi:10.1016/j.bushor.2019.11.006 [2022. 09. 30.]

Klein Tamás – Tóth András (szerk.) 2018: *Technológia jog – robotjog – cyberjog* [sic!]. Budapest: Wolters Kluwer.

Miskolczi Barna – Szathmáry Zoltán 2018: *Büntetőjogi kérdések az információk korában*. Budapest: HVG Orac.

Péterfalvi Attila – Révész Balázs – Buzás Péter (szerk.) 2021: *Magyarázat a GDPR-ról*. Budapest: Wolters Kluwer.

Schiff Beáta 2020: *A gépi tanulás és MI adatvédelmi kérdései a deepfake-technológia vonatkozásában*. Szakdolgozat, kézirat. Budapest: KRE ÁJK.

Szmolenszki Ildikó 2018: *A személyes adatok büntetőjogi védelme*. Szakdolgozat, kézirat. Budapest: ELTE ÁJK JTI.

Tóth Dávid 2020: Személyiséglopás az interneten. *Büntetőjogi Szemle*, 1. https://ujbtk.hu/wp-content/uploads/lapszam/BJSz_202001_113-119o_TothDavid.pdf [2022. 09. 30.]

FORRÁSOK

Datatilsynet 2018: *Artificial intelligence and privacy*. Report, January 2018. https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf [2022. 09. 29.]

Goble, Rachel 1999: The execution of Perkin Warbeck. *History Today*, november 11. https://www.historytoday.com/rachel-goble/execution-perkin-warbeck [2022. 09. 30.]

Hlács Ferenc 2016: AI: a nem emberi intelligencia már velünk van? 1. rész. *HWSW*, június 20. https://www.hwsw.hu/hirek/55760/ai-mesterseges-intelligencia-gepi-tanulas-machine-deep-learning.html [2022. 09. 30.]

Moyer, Edward 2022: Amazon, Google, Meta Among Targets of EU Law on Disinformation, Harmful Content. *CNET*, április 22. https://www.cnet.com/news/politics/amazon-google-meta-among-targets-of-eu-law-on-disinformation-harmful-content/ [2022. 09. 30.]

Nemzeti Adatvédelmi és Információszabadság Hatóság (NAIH) 2022: NAIH-305-5/2022. számú határozat. https://naih.hu/hatarozatok-vegzesek/file/532-szomszed-kamerasugy [2022. 09. 30.]

W1 = *Lambert Simnel English pretender*. https://britannica.com/biography/Lambert-Simnel-English-pretender [2022. 09. 30.]

W2 = https://www.youtube.com/watch?v=cQ54GDm1eL0 [2022. 09. 30.]

W3 = https://www.youtube.com/watch?v=enr78tJkTLE [2022. 09. 30.]

W4 = A Európai Parlament és a Tanács (EU) 2016/679. rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet). https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=celex:32016R0679 [2022. 09. 30.]

W5 = Az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról (javaslat), COM(2021) 206 final, Brüsszel, 2021. április 21. https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:52021PC0206&from=FR [2022. 09. 30.]

W6 = Az Európai Parlament és a Tanács rendelete a digitális szolgáltatások egységes piacáról (digitális szolgáltatásokról szóló jogszabály) és a 2000/31/EK irányelv módosításáról (javaslat), COM(2020) 825 final, Brüsszel, 2020. december 15. https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:52020PC0825&from=en [2022. 09. 30.]

W7 = Az Európai Parlament és a Tanács 2000/31/EK irányelve (2000. június 8.) a belső piacon az információs társadalommal összefüggő szolgáltatások, különösen az elektronikus kereskedelem egyes jogi vonatkozásairól (Elektronikus kereskedelemről szóló irányelv). https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=celex:32000L0031 [2022. 09. 30.]

W8 = Európai Bizottság: The Digital Services Act package. https://digital-strategy.ec.europa.eu/hu/node/27 [2022. 09. 30.]

MIKLÓS GELLÉRT

A deepfake-tartalmak szabályozása az Európai Unió jogában

Az Európai Unió a tagállami demokráciák, valamint a saját létezésére nézve az egyik legnagyobb kihívásként többek között az álhíreket és a jogellenes deepfake-tartalmakat jelölte meg. A mesterséges intelligencia robbanásszerű fejlődése által új technológiák kerültek kifejlesztésre, amelyekkel minden korábbinál élethűbb audiovizuális tartalmak hozhatók létre. Az EU ezért új jogszabályokat fogadott el, valamint régi jogszabályokat módosított az egységes belső piac harmonizációját, valamint az alapjogok védelmét és az innováción alapuló adatgazdaság kialakulását elősegítő jogszabályi keretrendszer megalkotása érdekében. Az alapjogok és társadalmi érdekek eredményesebb védelme érdekében kiszélesítésre került az online platform szolgáltatók kötelezettségeinek köre, és szigorú átláthatósági szabályok kerültek megállapításra. A különböző jogszabályok a jogellenes deepfake-tartalmak létrehozásának és terjesztésének különböző szakaszait és különböző szereplőit szabályozzák.

Kulcsszavak: mesterséges intelligencia, Európai Unió, online platformok szabályozása, GDPR

1. BEVEZETÉS

A közösségi média megjelenése gyökeresen változtatta meg a média működését, valamint emberek nagy részének hír- és tartalomfogyasztási szokásait. A felhasználók a közösségimédia-felületeken már nemcsak fogyasztják a híreket és a tartalmakat, hanem azok szerkesztőivé is válnak (Veszelszki 2021: 5). A közösségimédia-felületek hátterében dolgozó algoritmusok pedig gondoskodnak arról, hogy a felhasználók számára relevánsnak vélt tartalmak megjelenítésével maximalizálják a felületeken eltöltött időt. Az álhírek, a propaganda nem tekinthető új keletű jelenségnek, amint a képek számítógépes szerkesztése is évtizedek óta létező technológia. A mesterséges intelligencia (MI, angolul Artificial Intelligence, AI) robbanásszerű fejlődésével azonban olyan eljárások és alkalmazások jöttek létre, amelyek minden korábbinál valósághűbb hamis audiovizuális tartalmak létrehozását teszik lehetővé. Ezeket a tartalmakat deepfake-nek nevezik, egyszerre utal-

va az azt létrehozó mélytanulásos technológiára, valamint a videó hamis voltára. A fejezetben a deepfake-en olyan mesterséges intelligencia alkalmazásával, gépi és mélytanulásos eljárásokkal létrehozott manipulált vagy szintetikus hang- vagy képfelvételt értünk, amelyen olyan emberek láthatók vagy hallhatók, akik olyan dolgot tesznek vagy mondanak, amelyet soha nem tettek vagy mondtak (Huijstee et al. 2022: 5).

A deepfake-tartalmak létrehozását több technológia területén elért egyidejű áttörés teszi lehetővé. Egyrészt az algoritmusok képek alapján be tudják azonosítani az adott személyre jellemző azonosítási pontokat – ilyenek a szemöldök, az orr vagy a száj – és azok alapján az adott személyt felismerni. Ezzel párhuzamosan egyre több kép és videófelvétel érhető el az interneten a közösségimédia-felületeken vagy videómegosztó oldalakon ingyenesen hozzáférhető módon, ami jelentős adatbázist tesz elérhetővé az algoritmusok részére. A harmadik tényező, amely hozzájárul a hamisított videók minőségének javulásához, a hamisított videók szűrése terén elért egyre pontosabb megoldások megjelenése (Huijstee et al. 2022: 7). Paradox módon a kifinomultabb hamisítványfelismerés hozzájárul az ultrarealisztikus deepfake-videók létrehozásához. A folyamat során a mesterséges intelligencia két egymással szemben álló rendszere (az úgynevezett generatív ellenséges hálózatok) a rendelkezésére bocsátott nagy mennyiségű, megjelölt audiovizuális adat alapján feltérképezi a személyre jellemző azonosítási pontokat, majd létrehoz egy új képet, amit aztán egy algoritmus megvizsgál, annak megállapítása érdekében, hogy a létrehozott kép valós-e vagy hamis (Europol 2022: 8). Ha a képet az algoritmus hamisnak találja, akkor a program ezt feljegyzi, és újabb képet készít, építve a korábbi vizsgálatok eredményére. A folyamat addig zajlik, amíg az algoritmus már nem ismeri fel a kép hamis voltát (Westerlund 2019: 3). A másik lényeges technológia a hangklónozás, mely során egy program segítségével egy néhány perces beszédminta alapján szintetikus másolatot készítenek az ember hangjából, amely alapján aztán a program már képes lesz a kért szöveget a hangalany hangján felolvasni. A hangklónozó alkalmazás a megtévesztésig egyező eredmény elérése érdekében pontosan reprodukálja a hangalany hangszínét, beszéde ütemét, a hangmagasságot, sőt, még az illető légzését is, és képes ezeket a kért érzelemnek megfelelően alakítani (Vaccari–Chadwick 2020: 2). Szövegszintézis alkalmazásával egy program pedig képes az alanyra jellemző beszédekből és szövegekből alkotott tanító adatbázis alapján egy személyre jellemző új szöveget létrehozni, amely stílusát tekintve megegyezik a választott személy írásban és beszédben használt stílusával (Huijstee et al. 2022: 13).

2. A DEEPFAKE SZABÁLYOZÁSA AZ EURÓPAI UNIÓBAN

Tekintettel az Európai Unió (EU) struktúrájára és jogszabályi hierarchiájára, jelenleg egy többszintű szabályozás van kialakulóban, amely egyaránt magában foglalja az Európai Unió tagállamaiban közvetlenül alkalmazandó rendeleteit, a tagállami implementációhoz kötött irányelveket, valamint a tagállami alkotmányos normákat, polgári jogi és büntetőjogi normákat, valamint uniós és tagállami szabályozó hatósági iránymutatásokat, kötelező és önkéntes szakmai etikai kódexeket.

2.1. Az EU mesterséges intelligenciáról szóló rendelettervezete

Az Európai Unió Bizottsága 2021-ben készült el az Európai Unió belső piacán forgalomba hozott vagy elérhetővé tett mesterséges intelligencia szabályozására vonatkozó jogszabálytervezettel. A szabályozás kifejezett célja, hogy az Európai Uniót globális vezető szereplővé tegye a biztonságos, megbízható és etikus mesterséges intelligencia fejlesztésében.

A mesterségesintelligencia-rendelet tárgyi hatálya kiterjed az MI-rendszerek Európai Unión belüli forgalomba hozatalára, üzembe helyezésére és használatára – beleértve a deepfake-tartalmakat létrehozó MI-programokat is. A tervezet meghatározza a mesterségesintelligencia-gyakorlatokra vonatkozó tilalmakat, a nagy kockázatú MI-rendszerekre vonatkozó különös követelményeket, az ilyen rendszerek üzemeltetőire vonatkozó kötelezettségeket, valamint harmonizált átláthatósági szabályokat a kép-, audió- vagy videótartalom előállítására vagy manipulálására használt MI-rendszerek tekintetében.

Az MI-rendelet személyi hatálya kiterjed az Európai Unióban MI-rendszert forgalomba hozó vagy üzembe helyező szolgáltatókra, függetlenül ezen szolgáltatók letelepedési helyétől, az MI-rendszerek EU-n belüli felhasználóira, valamint az MI-rendszerek harmadik országban található szolgáltatóira és felhasználóira, ha a rendszer által előállított kimenetet az Unióban használják. A kimenet *(output)* a rendelet megfogalmazása alapján a felhasználó által meghatározott célkitűzésre az MI-rendszer által adott eredmény. Ilyen eredmény lehet az MI által generált tartalom, előrejelzés, ajánlás vagy döntés, például egy meghatározott feltételek alapján összeállított szöveg vagy következtetés. Tehát az MI-rendelettervezet szabályai kiterjednek akár egy harmadik országbeli MI-rendszer-felhasználóra is, amennyiben ismert EU-s politikusok képmását felhasználva gyűlöletbeszédet terjesztő jogellenes deepfake-tartalmat terjeszt az EU-ban.

Az MI-rendelet a mesterségesintelligencia-rendszereket kockázatuk alapján osztja fel: 1. tiltott, 2. nagy kockázatú, 3. korlátozott kockázatú, valamint 4. minimális kockázatú vagy nem kockázatos rendszerekre. Az Európai Bizottság tájékoztatása alapján az EU-ban jelenleg alkalmazott MI-rendszerek nagy része a

minimális kockázatú vagy nem kockázatos rendszerek közé tartozik (Európai Bizottság 2022). Az MI-rendelet tételes felsorolást nyújt a tiltott rendszerekről, ilyen például bármely olyan MI-rendszer, amely hatóságok által végzett közösségi pontozáshoz vezethet. A rendelet III. melléklete felsorolja azokat a nevesített példákat, amelyeket mindenképpen magas kockázatúnak kell tekinteni. Ebben a felsorolásban a deepfake-tartalmak létrehozásához szükséges MI-rendszerek nem szerepelnek, azonban a listán helyet kaptak a bűnüldöző hatóságok által a deepfake-tartalmak felderítésére szolgáló MI-rendszerek. Az Európai Unió egységes piacán történő forgalomba hozatal előtt a magas kockázatú rendszereknek számos követelménynek kell megfelelnie az MI-rendelet alapján. A deepfake-tartalmak létrehozásához szükséges MI-rendszerek korlátozott kockázatú rendszereknek minősülnek, amelyeknek ezáltal nem kell megfelelniük a magas kockázatú rendszerekkel szemben támasztott szigorú követelményeknek, alkalmazandók azonban a rendelet által előírt átláthatósági követelmények. Az MI-rendelet alapján azon MI-rendszerek felhasználói, amelyek olyan, meglévő személyekre, tárgyakra, helyekre vagy más szervezetekre vagy eseményekre érzékelhetően hasonlító képet, audió- vagy videótartalmat generálnak vagy manipulálnak, amely egy személy számára megtévesztő módon eredetinek vagy valóságosnak tűnhet, kötelesek azt közölni vagy a tartalmat oly módon jelölni, amelyből megállapítható, hogy a tartalmat mesterségesen hozták létre vagy manipulálták. Ez alól a követelmény alól kivételt képeznek azok az esetek, amikor a felhasználást törvény engedélyezi, vagy ha az a véleménynyilvánítás szabadságához, továbbá a művészet és tudomány szabadságához való jog gyakorlásához szükséges, és a harmadik felek jogaira és szabadságaira vonatkozó megfelelő biztosítékok hatálya alá tartozik. A fentiek alapján tehát nem minősül az MI-rendelet szabályaival ellentétesnek, ha egy Európai Unión kívül harmadik országban, például Kínában vagy Oroszországban található felhasználók a művészet vagy a véleménynyilvánítás szabadságának keretében ismert politikusok képmását felhasználva készítik el a deepfake-tartalmat, amennyiben azokat megfelelő jelöléssel látják el. Az MI-rendelet nem határozza meg konkrétan a közlés formai követelményeit, és arra vonatkozóan nem nyújt iránymutatást sem a felhasználók részére. Ebből kifolyólag a gyakorlat további iránymutatás nélkül valószínűleg számos eltérő megközelítést fog majd eredményezni.

Az MI-rendelet a szabályok megsértése esetére többszintű szankciórendszert állapít meg, amely a legsúlyosabb esetekben harmincmillió euró összegű vagy a vállalkozások előző pénzügyi év teljes éves világpiaci forgalmának legfeljebb 6%-át kitevő összegű közigazgatási bírság lehet. Az MI-rendeletben meghatározott egyéb követelmények és kötelezettségek megsértése esetén a közigazgatási bírság összege húszmillió euró vagy vállalkozások előző pénzügyi év teljes éves világpiaci forgalmának legfeljebb 4%-át kitevő összegű. Tekintettel arra, hogy korlátozott kockázatú MI-rendszerekre előírt átláthatósági kötelezettségek megsértésére az

MI-rendelet nem nevesít egyéb szankciót, így azok megsértésére kiszabható bírság maximuma elérheti a húszmillió eurót.

A fentiek alapján tehát az MI-rendszer felhasználója felelős az átláthatósági követelmények betartásáért, azok megszabása esetén a felhasználó felelősségre vonható, bírságolható. Ez egyben azt is jelenti, hogy a nem megfelelően jelölt deepfake-tartalmak miatt sem az azok létrehozását lehetővé tevő MI-rendszereket létrehozó szolgáltató, sem az importőr, sem a forgalmazó nem vonható felelősségre. Nem tartozik továbbá a rendelet hatálya alá és a rendelet alapján nem szankcionálható az sem, aki a felhasználó által az MI-rendszer alkalmazásával létrehozott, a követelményeknek nem megfelelően jelölt vagy nem jelölt deepfake-tartalmakat másokkal megosztja, azokat terjeszti. Kérdéses, hogy a fenti szankciórendszer az eleve ártó szándékkal létrehozott illegális tartalmak létrehozóival szemben mekkora visszatartó erőt jelenthet. Itt elég a hírességek arcképét felhasználó pornográf deepfake-tartalmakat internetes fórumokon anonim módon közzétevőkre vagy a deepfake-tartalmakat közéleti beavatkozás céljából létrehozókra gondolni. Az amerikai elnökválasztásba történő külföldi beavatkozások vizsgálata feltárta, hogy az álhírek nagy része harmadik országokból érkezett, amelyek egy részét anyagi megfontolásból magánszemélyek vagy azok egy csoportja, míg más részét államilag támogatott, specializált csoportok hozták létre és terjesztették (Chesney–Citron 2018: 1757). Egyelőre kérdéses, hogy az MI-rendeletben foglalt szankciók önmagukban milyen visszatartó erővel rendelkeznek az eleve ártó szándékú, sok esetben rejtve maradó felhasználókkal szemben.

2.2. Az általános adatvédelmi rendelet (GDPR)

A deepfake-tartalmak létrehozásához jellemzően szükséges valamely személyes adat felhasználása. A személyes adatok kezelését az Európai Unióban harmonizáltan az általános adatvédelmi rendelet (GDPR) szabályozza. A rendelet alapján személyes adatnak minősül az azonosított vagy azonosítható természetes személyre, azaz az érintettre vonatkozó bármely információ; például a természetes személy testi, fiziológiai azonosságára vonatkozó adatok. Ez azt jelenti, hogy az MI-rendszer tanulásához felhasznált minta részét képező, beszédet tartalmazó hangfelvételek vagy személyeket beazonosítható módon ábrázoló fotók vagy videók személyes adatnak minősülnek, éppúgy, mint a valós személyt beazonosítható módon ábrázoló deepfake-tartalmak is. Ez azt jelenti, hogy amennyiben a deepfake-tartalmat létrehozó programot megalkotó, EU-ban tevékenységi hellyel rendelkező szolgáltató – amennyiben a program a szolgáltató részére hozzáférést biztosít a személyes adatokhoz –, valamint az azt felhasználó személy által végzett adatkezelésre az általános adatvédelmi rendelet szabályai irányadók. Tekintettel arra, hogy a személyes adatok védelméhez fűződő jog minden magánszemélyt

megillet, ezért személyes adatot kezelni csak az általános adatvédelmi rendeletben meghatározott elvek és szabályok mentén, valamely jogalap megléte esetén lehet.

A deepfake-tartalmak létrehozásához szükséges jogalapot nagy valószínűséggel az érintett hozzájárulása, vagy az adatkezelő, vagy harmadik fél jogos érdekeinek érvényesítése fogja szolgáltatni. Utóbbira az adatkezelőnek akkor van lehetősége, amennyiben az érintett, tehát a deepfake-tartalom által felismerhetően ábrázolt személy jogai nem élveznek elsőbbséget a deepfake-tartalom létrehozójának alapvető jogaival szemben. Ilyen jogos érdek alapján végzett deepfake-tartalom létrehozására lehet példa egy közismert politikus parodisztikus megjelenítése egy mesterségesen létrehozott videóban. Ebben az esetben az alkotó véleménynyilvánítás szabadságához fűződő joga elsőbbséget élvez a politikus személyes adatainak védelméhez fűződő jogával. Az illegális célok elérésére szolgáló deepfake-tartalmak alkotói azonban nélkülözik a jogalapot az adatok jogszerű kezelésére – hiszen valószínűtlen, hogy rendelkezzenek a videón ábrázolt személyek előzetes beleegyezésével, vagy hogy jogos érdekük elsőbbséget élvezzen az érintettek jogaival szemben –, ezért ezen deepfake-tartalmak mindenképpen jogszerűtlen adatkezelést valósítanak meg.

Az általános adatvédelmi rendelet az érintetteket jogaik hatékony védelme érdekében különböző jogokkal ruházza fel. Az érintettet egyrészt tájékoztatni kell személyes adatai kezeléséről, másrészt hozzáférési joggal rendelkezik az adatkezeléssel kapcsolatos bizonyos információkhoz. Meghatározott esetekben – például ha a személyes adatok kezelése jogellenesen történt, akkor – az érintett követelheti a rá vonatkozó személyes adatok törlését. Ez azt jelenti, hogy amennyiben egy személy képmását hozzájárulása nélkül használják fel egy deepfake-tartalom létrehozására, és az alkotó oldalán nem került megjelölésre olyan érdek, amely elsőbbséget élvezne az érintett jogaival szemben, akkor az érintett kérésére a deepfake-tartalmat törölni kell. Tekintettel arra, hogy a deepfake-tartalmak jelentős része ismeretlen személyek által illegális cél megvalósítására kerül létrehozásra, így a gyakorlatban az érintett törléshez való jogának érvényesítése akadályokba ütközik. Egy átlagos személy nem rendelkezik sem megfelelő szakértelemmel, sem erőforrásokkal ahhoz, hogy felderítse egy deepfake-tartalom ismeretlen létrehozójának kilétét annak érdekében, hogy jogait vele szemben gyakorolhassa. Ebben az esetben is nyitva áll a lehetőség az érintett előtt, hogy az érintett panaszt nyújtson be a felügyeleti hatóságnál – amely Magyarországon a Nemzeti Adatvédelmi és Információszabadság Hatóság (NAIH) – személyes adatai jogellenes kezelésre történő hivatkozással. A hatóság köteles a panasz tárgyát a szükséges mértékben kivizsgálni és a vizsgálat eredményéről a panasztevőt tájékoztatni.

Látható, hogy a jogérvényesítés eredményességét két tényező nagyban befolyásolja: egyrészt az érintettek adatvédelmi tudatossága és jogismerete, az, hogy mennyire ismerik személyes adataik védelméhez fűződő jogaikat és azok érvényesítésének eszközeit, másrészt pedig az, hogy az adatvédelmi hatóság milyen eszkö-

zökkel rendelkezik a jogellenes adatkezelésekkel kapcsolatos panaszok kivizsgálására, felderítésére. Amennyiben a jogellenes deepfake-tartalom létrehozójának kiléte megállapítható, és az érintett vagyoni vagy nem vagyoni kárt szenvedett, úgy az általa elszenvedett kárért az adatkezelőtől, azaz a tartalom létrehozójától kártérítést követelhet, az adatvédelmi hatóság pedig bírságot szab ki, amelynek legmagasabb összege húszmillió euró vagy a vállalkozások előző pénzügyi éve teljes éves világpiaci forgalmának legfeljebb 4%-át kitevő összeg lehet. Ezen jogvédelmi eszközök eredményessége azonban kérdéses ismeretlen személyazonosságú vagy harmadik országbeli adatkezelőkkel szemben.

2.3. Elektronikus kereskedelemről szóló irányelv

Az EU-n belül az online tartalmak egységes szabályozásának egyik első lépése volt az elektronikus kereskedelemről szóló 2000/31/EK irányelv elfogadása. Az irányelv célja a belső piac harmonizációja és az információs társadalommal összefüggő szolgáltatásokra vonatkozó szabályok közelítése volt. Az irányelvet huszonkét évvel ezelőtt fogadták el, egy olyan időszakban, amikor a ma ismert legnagyobb közösségimédia-felületek még nem léteztek, az internet emberek életére gyakorolt hatása, valamint a gazdaságban betöltött szerepe pedig összehasonlíthatatlanul kisebb volt napjainkban betöltött szerepénél. A 2015/1535 irányelv meghatározása alapján információs társadalommal összefüggő szolgáltatás az információs társadalom bármely szolgáltatása, azaz bármely, általában térítés ellenében, távolról, elektronikus úton és a szolgáltatást igénybe vevő egyéni kérelmére nyújtott szolgáltatás. Az elektronikus kereskedelemről szóló irányelv egységesítette a közvetítő szolgáltatók felelősségére vonatkozó szabályokat. A szabályok alapján a közvetítő szolgáltatók, azaz olyan szolgáltatók, amelyek a szolgáltatás igénybe vevője által küldött információnak hírközlő hálózaton keresztül történő továbbítását vagy hírközlő hálózathoz való hozzáférést biztosítanak, nem felelősek a továbbított információért, amennyiben nem a szolgáltató kezdeményezi az adatátvitelt, választja ki az adatátvitel címzettjét; és választja meg vagy módosítja a továbbított információt. Az irányelv az előbb ismertetett egyszerű továbbításra (angolul *mere conduit*) vonatkozó mentességi szabályokhoz hasonló szabályokat fogalmaz meg a gyorsítótárolóban történő rögzítés *(caching)*, valamint a tárhelyszolgáltatás nyújtása vonatkozásában.

Az irányelvben nem került meghatározásra a jogellenes tartalom fogalma, rögzítésre került viszont azt, hogy a tagállamok nem állapíthatnak meg olyan kötelező érvényű általános kötelezettséget a közvetítő szolgáltatóra nézve az egyszerű továbbítás, a caching, valamint a tárhelyszolgáltatás nyújtása során, amely azt az általa továbbított vagy tárolt információk nyomon követésére vagy jogellenes tevékenységre utaló tények vagy körülmények vizsgálatára kötelezné. A fentiek

alapján egy internetszolgáltató vagy egy felhőtárhely-szolgáltató nem felelős a meghatározott feltételek fennállása esetén a szolgáltatás felhasználásával továbbított vagy tárolt jogsértő tartalomért, akkor sem, ha az a tartalom jogellenes célból létrehozott deepfake-tartalom, feltéve azonban, hogy amint a jogellenes deepfake-tartalomról tudomást szerzett, haladéktalanul intézkedik az információ eltávolításáról vagy az ahhoz való hozzáférés megszüntetéséről. A tárhelyszolgáltatók felelősségének körére, valamint a törlésre kötelezésre és az általános nyomon követésre vonatkozó tilalom értelmezésére vonatkozóan releváns az Európai Unió Bírósága a 2019. október 3-i Glawischnig–Piesczek-ügyben (C-18/18) hozott ítélete. Az idézett ügyben a felperes osztrák politikus, az osztrák nemzeti tanács tagja kérte a Facebookot, hogy töröljön egy bármely felhasználó által látható becsületsértő hozzászólást egy, a politikus arcképével illusztrált, menekültekkel kapcsolatos cikk alól. Mivel a Facebook a kérdéses hozzászólást nem távolította el, ezért a felperes végül bírósághoz fordult a Facebook-hozzászólás törlésére kötelezése érdekében. Az eljárás során a legfelsőbb bíróság előzetes döntéshozatali eljárás keretében az Európai Unió Bíróságához fordult, tekintettel arra, hogy a jogvita az uniós jog értelmezésével kapcsolatos kérdéseket vetett fel. A bírósági eljárás során nem volt vitatott, hogy a Facebook a közösségimédia-szolgáltatása nyújtása során tárhelyszolgáltatóként is eljár. A bíróság az ügyben végül három fontos megállapítást is tett: 1. Egyrészt a fentebb ismertetett általános nyomon követési kötelezettség hiánya nem zárja ki azt, hogy valamely tagállami bíróság attól függetlenül kötelezze a tárhelyszolgáltatót az általa tárolt és a korábban jogellenesnek nyilvánított információval azonos tartalmú információk törlésére vagy az azokhoz való hozzáférés megakadályozására, hogy ki kérelmezte ezen információk tárolását. 2. Továbbá azt sem, hogy kötelezze a tárhelyszolgáltatót az általa tárolt és a korábban jogellenesnek nyilvánított információval azonos értelmű információk törlésére vagy az azokhoz való hozzáférés megakadályozására, amennyiben az érintett információk ezen eltiltásból eredő nyomon követése és vizsgálata az olyan üzenetet közvetítő információkra korlátozódik, amelynek tartalma lényegében változatlan a jogellenesség megállapításának alapjául szolgáló üzenetétől, és amely tartalmazza a jogsértéstől való eltiltásban meghatározott elemeket, továbbá amennyiben az ezen azonos értelmű megfogalmazás tekintetében a korábban jogellenesnek nyilvánított információt jellemző megfogalmazáshoz képest fennálló különbségek nem követelik meg a tárhelyszolgáltatótól e tartalom önálló értékelését. 3. Mindezek mellett, hogy releváns nemzetközi jogi kereteken belül világszinten kötelezze a tárhelyszolgáltatót a jogsértéstől való eltiltással érintett információk törlésére vagy az azokhoz való hozzáférés megakadályozására. Ez azt jelenti, hogy egy érintett, akivel kapcsolatban személyes adatai jogellenes kezelésével készült deepfake-tartalom feltöltésre kerül egy közösségimédia-felületre, kérheti ennek a jogellenes tartalomnak az eltávolítását, és amennyiben a szolgáltató a kérésnek nem tesz eleget, úgy nemzeti bíróság előtt követelheti a jogellenes

tartalom világszintű törlését, valamint minden azzal egyező megosztás vagy újraközzététel megakadályozását is. Ez ugyan hatékony megoldásnak tűnhet, azonban a gyakorlatban a tudomásszerzéstől a szolgáltató által nem teljesítés esetén a bíróság jogerős ítéletéig hetek, hónapok telhetnek el, ami a közösségi média világában beláthatatlanul hosszú idő.

Az internetszolgáltatók esetében a 2002/58/EK elektronikus hírközlési adatvédelmi irányelv rögzíti a közlések titkosságának követelményét, amely alapján a nyilvános hírközlő hálózatok és a nyilvánosan elérhető elektronikus hírközlési szolgáltatások segítségével történő közlések és az azokra vonatkozó forgalmi adatok titkosak, a közlések és az azokra vonatkozó forgalmi adatok felhasználókon kívüli személyek által történő, az érintett felhasználó hozzájárulása nélküli meghallgatása, lehallgatása, tárolása vagy más módon történő elfogása vagy megfigyelése kifejezetten tilos. Tehát az internetszolgáltatók a nemzeti jogszabályokban meghatározott szűk körű kivételektől eltekintve nem figyelhetik vagy szűrhetik a hálózaton továbbított információkat jogellenes tartalom után kutatva. Ez egy nagyon fontos alapjogi korlátozás, amely egyben azonban azt is jelenti, hogy a jogellenes deepfake-tartalmak terjesztését az internetszolgáltatók nem szűrhetik, így nem is korlátozhatják. Az elektronikus kereskedelemről szóló irányelv fenti, felelősségi rendelkezéseivel párhuzamosan a tagállami jogszabályok szabályozzák a szolgáltató jogsértés megszüntetésére vagy megelőzésére kötelezésének, valamint a tárhelyszolgáltatók vonatkozásában az információ eltávolításának vagy a hozzáférés megszüntetésének szabályait.

2.4. Digitális szolgáltatásokról szóló jogszabály (Digital Services Act)

Tekintettel arra, hogy a jelenleg még hatályban lévő, elektronikus kereskedelemről szóló irányelv szabályai a technológiai fejlődés következtében mára már nagyrészt elavulttá váltak, valamint arra, hogy az irányelvi szintű szabályozás az eltérő tagállami végrehajtás következtében az Európai Unió egységes belső piacán eltérő szabályozást eredményezett, szükségessé vált a szabályok felülvizsgálata. A digitális szolgáltatásokról szóló rendelet egy lépcsőzetes szabályozási rendszert vezet be, amelyben a szolgáltatók típusától és nagyságától függ a jogszabályi követelmények száma. A digitális szolgáltatásokról szóló rendelet hatálya a közvetítő szolgáltatókra, az egyszerű továbbítást, a gyorsítótárazást vagy tárhelyet biztosító szolgáltatókra terjed ki, tekintettel arra, hogy szerepük az ilyen szolgáltatások exponenciális igénybevétele miatt jelentősen megnőtt a jogellenes vagy egyéb módon káros információk és tevékenységek közvetítésében és terjesztésében.

A rendelet a közvetítő szolgáltatók székhelyétől vagy lakóhelyétől függetlenül alkalmazandó, amennyiben szolgáltatásokat nyújtanak az Unióban. Egy székhellyel nem rendelkező szolgáltató esetében megállapítható az érdemi kapcsolat az

Európai Unióval, amennyiben a közvetítő szolgáltató egy vagy több tagállamban jelentős számú felhasználóval rendelkezik, vagy a tevékenységei egy vagy több tagállam felé irányulnak. A fentiek alapján az Európában elérhető és népszerű közösségimédia-szolgáltatók a rendelet hatálya alá fognak tartozni, mint a Facebook vagy a Twitter. A rendelet a tárhelyszolgáltatók csoportján belül elkülöníti az online platformok alkategóriáját, amelybe például a közösségi hálózatok és az online piacterek tartoznak bele.

A rendelet pótolja továbbá az elektronikus kereskedelemről szóló irányelv egyik hiányosságát, és meghatározza a jogellenes tartalom fogalmát, amelyen bármely olyan információ értendő, amely önmagában vagy egy tevékenységre való hivatkozással, beleértve a termékek értékesítését vagy a szolgáltatások nyújtását, nem felel meg az uniós jognak vagy valamely tagállam jogának, függetlenül az adott jog pontos tárgyától vagy jellegétől. A 12. Preambulumbekezdés számos olyan példát is felsorol a jogellenes tartalomra, amely deepfake-tartalom létrehozásával is megvalósítható, ilyen például a jogellenes gyűlöletbeszéd vagy a terroristatartalom és a jogellenes megkülönböztető tartalom, vagy jogellenes tevékenységekkel kapcsolatos, például a gyermekek szexuális zaklatását ábrázoló képek megosztása, magánjellegű képek jogellenes, nem beleegyezésen alapuló megosztása, online zaklatás vagy a szerzői joggal védett anyagok nem engedélyezett felhasználása. A rendelet fenntartja a korábbi irányelvben rögzített felelősség alóli mentességi szabályokat, valamint azt is rögzíti, hogy a közvetítő szolgáltatóktól nem tagadható meg a felelősség alóli mentesség kizárólag azon az alapon, hogy önkéntes, saját kezdeményezésű vizsgálatokat vagy egyéb, a jogellenes tartalom észlelésére, azonosítására és eltávolítására, illetve a hozzáférés megszüntetésére irányuló tevékenységet folytatnak.

Továbbra sem kerül azonban előírásra aktív nyomon követési kötelezettség, amely kötelezné a szolgáltatókat a továbbított vagy tárolt információk követésére vagy a jogellenes tevékenységre utaló tények vagy körülmények aktív feltárására. A fentiek alapján az egyszerű továbbítást végző internetszolgáltatók továbbra sem vizsgálhatják a továbbított információ jogellenességét, de a közösségimédia-felületek szolgáltatói számára azonban adott a lehetőség. A rendelet rögzíti a fentebb idézett európai bírósági döntésben foglaltakat, és rendelkezik a jogellenes tartalom elleni fellépésre vonatkozó határozatok végrehajtásáról. A határozatok gyorsabb végrehajtásának elősegítése és a hatékonyabb tartalommoderálás érdekében a rendelet új kötelezettségként bevezeti az átláthatósági jelentési kötelezettséget, amely kötelezi a közvetítő szolgáltatókat arra, hogy legalább évente egyszer egyértelmű, könnyen érthető és részletes jelentést tegyenek közzé az előző jelentés óta eltelt időszakban végzett tartalommoderálásról. A jelentésnek tartalmaznia kell a (i) tagállami hatóságok által kibocsátott határozatok számát a jogellenes tartalom típusa szerint, valamint a határozatokban meghatározott cselekvéshez szükséges átlagos időt, (ii) a személyek vagy szervezetek által tett jogellenes tartalomra vo-

natkozó bejelentések számát az állítólagos jogellenes tartalom típusa szerint, (iii) a saját kezdeményezésére történt tartalommoderálásra vonatkozó információkat, valamint (iv) belső panaszkezelési rendszerben beérkezett panaszok számát és az azokra vonatkozó részletes adatokat. Az online platformokat ezen túlmenően kötelezi arra, hogy szolgáltassanak információt a (v) peren kívüli vitarendezési testületek elé terjesztett viták számáról, (vi) a visszaélések miatti felfüggesztések számáról, beleértve a jogellenes tartalom megosztása miatti felfüggesztést is, valamint a (vii) tartalommoderálás céljából használt automatizált eszközökről. Ez egy nagyon jelentős előrelépés a tárhelyszolgáltatók és különösen az online platformok tartalommoderálásának átláthatósága terén, hiszen az adatok nyilvánossága mellett adott esetben komoly reputációs kockázatot fog jelenteni a szolgáltatóknak a hatósági határozatok nem teljesítése vagy a jogellenes tartalmak lassú törlése.

A jogellenes tartalom elleni fellépés gyorsabbá és megbízhatóbbá tétele érdekében a rendelet megalkotta a megbízható bejelentő fogalmát. Az ilyen megbízható bejelentők csak közjogi vagy nem kormányzati szervezetek lehetnek, amelyek a kollektív érdekek képviseletében járnak el a jogellenes tartalmak felügyelete során és rendelkeznek az ilyen tartalom kezeléséhez szükséges szakértelemmel. Az online platformok kötelesek a megbízható bejelentőktől érkező bejelentéseket kiemelten és késedelem nélkül kezelni.

Szintén új követelmény az online platformokkal szemben, hogy a rendszeresen jogellenes tartalmat megosztó felhasználók számára fel kell függeszteniük szolgáltatásuk nyújtását. A szolgáltatás nyújtásának felfüggesztése valószínűleg azokkal a felhasználókkal szemben lesz hatásos, akik a jogellenes tartalmakat, beleértve a deepfake-tartalmakat is meggyőződésből vagy a megfelelő edukáció hiányából fakadóan osztják meg a közösségimédia-felületen. Azokkal szemben, akik jogellenes cél megvalósítása érdekében hozzák létre a deepfake-tartalmakat, és azt anonim módon vagy álprofilok létrehozása útján terjesztik az online platformokon, ez az intézkedés várhatóan kevésbé lesz hatékony, hiszen nem gátolja meg új profil létrehozását.

A rendelet további szabályokat fogalmaz meg az online óriásplatformokra vonatkozóan. Amint az a rendelet az 54. Preambulumbekezdésben is kiemeli, az online óriásplatformok társadalmi kockázatokkal járhatnak, amelyek hatálya és hatása eltér a kisebb platformok kockázataitól. Attól kezdődően, hogy egy platform igénybe vevői az uniós népesség jelentős hányadát teszik ki, a platformból eredő rendszerszintű kockázatok aránytalanul negatív hatást gyakorolnak az Unióban. A rendelet az online óriásplatformok jelentette rendszerkockázatokat három csoportra osztja: (i) az első csoport a platformok nyújtotta szolgáltatással visszaélések, ilyenek például a jogellenes tartalom terjesztésével kapcsolatos kockázatok, (ii) a második kategória a szolgáltatás által alapvető jogok gyakorlására kifejtett hatások, (iii) a harmadik kategória a platformok szolgáltatásainak szándékos és gyakran koordinált manipulálása, ami előre látható hatással van többek között

a társadalmi párbeszédre, a választási folyamatokra, a közbiztonságra. A jogellenes deepfake-tartalmak mind az első, mind a harmadik rendszerszintű kockázatcsoportba beleilleszthetők. Az első kategória kapcsán az online óriásplatformok esetén jelentős rendszerszintű kockázatot jelent a jogellenes tartalom terjesztése, különösen a rendkívül nagy bázissal rendelkező fiókok útján. A harmadik kategóriára példa a hamis fiókok létrehozása, botok használata vagy egyéb automatizált vagy részben automatizált magatartások, amelyek a jogellenes tartalomnak minősülő vagy az online platform szerződési feltételeivel összeegyeztethetetlen információk gyors és széles körű terjedéséhez vezethetnek. Konkrét példa lehet egy tagállami választásba történő beavatkozási kísérlet, amely során harmadik államból szervezetten létrehozott hamis fiókok felhasználásával történik az álhírek, valamint a jogellenes deepfake-tartalmak terjesztése.

2.5. Audiovizuális médiaszolgáltatásokról szóló irányelv (2018/1808 irányelv)

Az Európai Unió audiovizuális médiaszolgáltatásokról szóló irányelve 2018-ban került módosításra, reagálva az audiovizuális tartalmak piacának terén végbement technológiai fejlődésre. A műszaki fejlődés új típusú szolgáltatásokat tett lehetővé, megváltoztatva a fiatalabb generációk nézői szokásait. A televízió veszített népszerűségéből, jelentősen megnőtt viszont a hordozható eszközökön elért videómegosztó platformokon keresztül audiovizuális tartalmakat fogyasztó nézők száma. A videómegosztó platformok népszerűségének megugrásával azonban az azon keresztül megjelenő káros tartalom aránya is megnőtt. Ennek egyik oka az, hogy a videómegosztó platformokon elérhető audiovizuális tartalom jelentős hányada nem tartozik a videómegosztó platform szolgáltatójának szerkesztői felelősségi körébe, hanem annak felhasználói töltik fel és teszik elérhetővé azokat. Az irányelv alapján a tagállamok kötelesek olyan szabályozást kialakítani, amely (i) megvédi a kiskorúakat az olyan audiovizuális tartalmaktól, amelyek károsíthatják a fizikai, szellemi vagy erkölcsi fejlődésüket, (ii) a közönséget az erőszakra vagy gyűlöletre uszító tartalmaktól, valamint (iii) azoktól a tartalmaktól, amelyek terjesztése terrorista bűncselekmény elkövetésére uszításnak, gyermekpornográfiának, rasszizmussal és idegengyűlölettel kapcsolatos bűncselekménynek minősül.

Az irányelv nem ír ugyan elő kötelező proaktív tartalomszűrési kötelezettséget a videómegosztó platformszolgáltatók részére, azonban biztosítaniuk kell annak lehetőségét, hogy a felhasználók jelezhessék a problémásnak, sértőnek, jogellenesnek tartott videókat, és bejelenthessék vagy megjelölhessék őket (Sorbán 2019: 215). Biztosítani kell továbbá a lehetőséget arra is, hogy a felhasználók jelezhessék, ha gyűlöletre uszító, kiskorúakra káros vagy bűncselekményt megvalósító tartalmat észlelnek. A platformoknak azonban nemcsak a kifogásolható tartal-

mak bejelentésének lehetőségét kell megteremteni, hanem biztosítani kell azt is, hogy a felhasználók bejelentéseit érdemi vizsgálat kövesse annak lehetőségével, hogy szükség esetén a meghatározott tartalmak törlésre kerüljenek vagy hozzáférhetetlenné váljanak bizonyos felhasználói csoportok számára. Az irányelv tehát rögzíti olyan intézkedéseket és a videómegosztóplatform-szolgáltatókra vonatkozó kötelezettségeket, amelyek alkalmasak a jogellenes deepfake-tartalmak – mint amilyenek például a pornográf deepfake-videók, az álhíreket vagy gyűlöletbeszédet terjesztő deepfake-videók – terjedésének visszaszorítására és terjedésük megakadályozására.

2.6. Az EU dezinformációval kapcsolatos megerősített magatartási kódexe (2022)

A korábban felsorolt jogalkotási aktusokkal ellentétben a dezinformációval kapcsolatos magatartási kódex egy iparági szereplők önkéntes alávetésén alapuló önszabályozási mechanizmus, amelyet az Európai Unió Bizottsága 2018-ban kezdeményezett, tekintettel arra, hogy a Bizottság az álhíreket és a hamis információkat jelölte meg a tagállami demokráciákra és az EU-ra irányuló egyik legnagyobb veszélyként. A Bizottság felülvizsgálata során számos hiányosságot azonosított, amelyek orvoslása és a hatékonyabb végrehajtás érdekében végül a kódex megerősítésére került sor. A kódex aláírói között megtalálhatók a legnagyobb iparági szereplők, mint a Microsoft, Meta, Google, TikTok, Twitter. A 2022. évi magatartási kódex az aláírók által végzett munka eredménye, az nem függ az Európai Unió Bizottságának jóváhagyásától. A kódex aláírói szabadon döntenek arról, hogy mely kötelezettségvállalásokat írják alá. Mivel a dezinformáció és a politikai propaganda hatékony eszközei lehetnek a deepfake-tartalmak, ezért a megerősített kódex is tartalmaz kifejezetten a deepfake-tartalmakkal kapcsolatos kötelezettségvállalást. Az aláíró felek a 14. számú kötelezettségvállalásban vállalják azt, hogy szolgáltatásuk nyújtása során fellépnek a hamis információkkal és a dezinformációval szemben, amelyek terjesztésének egyik módja jogellenes deepfake-tartalmak terjesztése. A 15. számú kötelezettségvállalás alapján a mesterségesintelligencia-rendszereket fejlesztő vagy üzemeltető aláírók, akik szolgáltatásaikon keresztül mesterséges intelligencia által generált és manipulált tartalmat terjesztenek – mint amilyenek a deepfake-tartalmak is –, kötelezettséget vállalnak arra, hogy figyelembe veszik tevékenységük során a mesterségesintelligencia-rendeletben foglalt átláthatósági kötelezettséget és a tiltott gyakorlatok jegyzékét.

3. ÖSSZEGZÉS

Amint az a fentiekből is látszik, az Európai Unió kiemelt fenyegetésként kezeli az álhíreket és a deepfake-tartalmakat. A deepfake-tartalmak létrehozásának különböző fázisait és szereplőit lehet megkülönböztetni: elkülöníthető egymástól a deepfake-tartalmat létrehozó személy, az azt lehetővé tevő technológia, a tartalom terjedése vagy terjesztése, az ábrázolt vagy imitált személy és a közönség (Huijstee et al. 2022: 10–12). Az európai uniós szabályozásról megállapítható, hogy több különböző fázisra is megfogalmaz szabályokat. Ez a megközelítés elengedhetetlen a hatékony fellépés érdekében. A szabályok és szankciók egy részének visszatartó ereje azonban így is kérdéses, tekintettel a deepfake-tartalmak létrehozásának dinamikájára. A szabályozás a jogellenes szándékkal létrehozott, anonim felhasználók által terjesztett deepfake-tartalmakkal szemben kevésbé hatékony, a digitális szolgáltatásokról szóló jogszabályban megfogalmazott intézkedések azonban eredményesek lehetnek a deepfake-tartalmak terjedésének visszaszorításában és ezáltal a negatív hatások széles körű csillapításában. A hatékony fellépéshez egy többdimenziós európai uniós szabályozás mellett szükség van egy azt kiegészítő, hatékony nemzeti szabályozásra is, valamint a személyek és a társadalom folyamatos edukációjára.

SZAKIRODALOM

Bateman, Jon 2020: Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenario. *Cybersecurity and the Financial System*, 7, július 8. https://carnegieendowment.org/files/Bateman_FinCyber_Deepfakes_final.pdf [2022. 09. 22.]

Chesney, Robert – Citron, Danielle K. 2018: Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107: 1753–1819. http://dx.doi.org/10.2139/ssrn.3213954

Europol 2022, Facing reality? Law enforcement and the challenge of deepfakes: An observatory report from the Europol innovation lab. *Publications Office of the European Union*, május 5., https://data.europa.eu/doi/10.2813/08370 [2022. 09. 25.]

Huijstee, M. – Boheemen, P. – Das, D., et al. 2022: Tackling deepfakes in European policy, *European Parliament, Directorate-General for Parliamentary Research Services, European Parliament*, https://data.europa.eu/doi/10.2861/325063 [2022. 09. 18.]

Sorbán Kinga 2019: A videómegosztóplatform-paradoxon, avagy az új európai szabályok alkalmazhatósága a globalizáció és az eltérő tagállami implementáció keresztmetszetében. *In Medias Res: folyóirat a sajtószabadságról és a médiaszabályozásról*, 8(2): 210–229.

Vaccari, C. – Chadwick, A. 2020: Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society* 6(1): 2. https://doi.org/10.1177/2056305120903408

Veszelszki Ágnes 2021: deepFAKEnews: Az információmanipuláció új módszerei. In: Balázs László (szerk.): *Digitális kommunikáció és tudatosság.* Budapest: Hungarovox Kiadó. 93–105.

Westerlund, Mika 2019: The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11): 40–53. http://doi.org/10.22215/timreview/1282

Európai Bizottság 2022: Regulatory framework proposal on artificial intelligence. *Digital Strategy website, European Commission,* augusztus 28. 2022. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#:~:text=The%20Regulatory%20Framework%20defines%204,Limited%20risk [2022. 09. 21.]

INFLUENSZEREK

GOCSÁL ÁKOS

Manipulált beszéd használata a személyészlelés kutatásában

A deepfake-technológia alkalmazásával valósnak tűnő virtuális személyek jönnek létre. Ahogy a valós világban, úgy a virtuális személyeknél is fontos tulajdonság a beszélő hangjának minősége, ugyanis a hang alapján a hallgatóban kialakult akusztikai élmény meghatározhatja a beszélő iránti attitűdöket, sőt, döntéseket is megalapozhat. Kutatói feladat annak megállapítása, hogy a beszéd akusztikai szerkezetében melyek azok a paraméterek, amelyek szerepet játszanak a beszélőről kialakított benyomásban. A jelen kutatásban 51 egyetemi hallgatónak 9-9 férfi és női bemondást játszottunk le, majd a hang kellemességéről és elképzelt döntési helyzetekben az attitűdjükről kérdeztük őket hétfokú skálák alkalmazásával. A beszédminták közül csak egy-egy volt természetes ejtésű, a többinek tempóján és alaphangján manipulációkat hajtottunk végre. Az eredmények azt az általános tendenciát mutatják, hogy a mélyített és gyorsított beszédminták inkább kedvező, az emelt és lassított bemondások inkább kedvezőtlenebb attitűdöt alakítottak ki. A kutatás a holdudvarhatást is kimutatta, a kellemesebb hangzású beszéd együtt járt az elképzelt beszélővel szembeni elfogadóbb attitűddel, kedvezőbb döntésekkel.

Kulcsszavak: beszéd, manipuláció, benyomás, naiv jellemzés, holdudvarhatás

1. BEVEZETÉS

A „deepfake" kifejezés valós személy illúzióját keltő videófelvételre, hangra vagy képre utal, amelyet részben vagy teljes egészében fejlett technológia alkalmazásával hoztak létre (Graber-Mitchell 2021). A deepfake létrehozásába tehát nemcsak a személy képi megjelenítése tartozik bele, hanem az is, ha a virtuális személy emberi hangon, adott esetben valamely célszemély hangján szólal meg. Összefoglaló néven hangklónozásnak nevezzük természetes beszédminták manipulációját, amelynek következtében a beszéd hallgatójában az a benyomás alakul ki, mintha az elhangzottakat ténylegesen az adott személy mondta volna ki. Khanhani, Watson és Janeja (2021) a deepfake beszéd azonosításával foglalkozó munkájukban annak három típusát különböztették meg. A visszajátszáson alapuló deepfake

valójában a célszemély előzetesen rögzített hangjának lejátszását jelenti. Ennek altípusa a kivágás-beillesztés módszer *(cut and paste)*, amely beszédrészletek összeillesztésével készített, az adott formában el nem hangzott mondatok előállítását jelenti. A deepfake-beszéd második nagy kategóriájához a gépi beszéd tartozik. Ebben az esetben egy szoftveres vagy hardveres szintetizátor a begépelt szöveget hangzó formába alakítja át. A harmadik nagy típust pedig a hangkonverzió vagy megszemélyesítés *(impersonation)* alkotja. Ekkor egy rendelkezésre álló beszédfelvétel módosítása történik oly módon, hogy az úgy hangozzon, mintha egy másik személy mondta volna. Megszemélyesítéskor nemcsak az alaphang és más akusztikai minőségek utánzása történik, hanem az így manipulált beszéd a célszemély beszédstílusát is meggyőzően tükrözi (Gao–Singh–Raj 2018). A hangkonverzió valójában a hangtranszformáció speciális esete. Transzformáció esetében általában nem cél más beszélő utánzása, hanem a beszédnek csak olyan módosítása történik, amely a nyelvi tartalmakat változatlanul hagyja, miközben annak hangzása természetes marad (Stylianou 2009). Itt említhetjük például a beszéd gyorsítását vagy lassítását, illetve a beszélő alaphangjának (hangmagasságának) emelését vagy mélyítését. Ezek a manipulációk bizonyos határokon belül megőrzik a beszéd természetes hangzását.

A virtuális személyek felépítésekor nem csak az a feladat a kutató, fejlesztő számára, hogy a legmeggyőzőbb, leginkább életszerű figurát kidolgozza. Emellett fel is kell mérnie, a befogadók, nézők, hallgatók hogyan vélekednek a virtuális személyről, milyen tulajdonságokkal ruházzák fel, mennyire találják elfogadhatónak. A természetes beszéd hangzása által kiváltott benyomásokkal már az 1920-as évek vége óta foglalkoznak kutatók. A nyelvész-antropológus Sapir (1927/1991) a beszéd egyénre jellemző sajátosságairól még elvi alapokon írt, nem sokkal később azonban már pszichológusok kísérleteket is végeztek ezzel kapcsolatban. A korai kísérletekben a hallgatóknak az elhangzott beszéd alapján a beszélő nemét, életkorát, intelligenciáját, foglalkozását, személyiségjegyeit kellett megítélniük (például: Pear 1931; Herzog 1933; Allport–Cantril 1934). Már ezekben a kísérletekben is megmutatkozott, hogy a hallgatók sztereotípiákat alkalmaznak a személyészlelés során: gyakoriak voltak a téves becslések, ennek ellenére meglehetősen nagy egyetértés mutatkozott a hallgatók között.

Némi kihagyás után az 1960-as években új lendületre kaptak a kutatások. Ezen a területen klasszikusnak tekinthető Zuckerman és Driver (1989) kísérlete, amelyben a szerzők egyrészt kimutatták, hogy a kísérleti személyek között egyetértés mutatkozott abban, hogy mely hangok tűntek számukra vonzónak vagy kevésbé vonzónak, másrészt kimutatták, hogy a vonzóbbnak ítélt hang együtt járt az elképzelt beszélő személyiségének kedvezőbb megítélésével. Weirich (2008) például azt találta, hogy a vonzóbb hangú beszélőt egyúttal jóindulatúbbnak is vélték a hallgatók. A hang minőségének és a vonzóságítéleteknek számos következményük van az emberi kapcsolatok alakulására. Suire és munkatársai (2021) a szakirodalomban közölt ered-

mények részletes áttekintése alapján megállapították, hogy a mélyebb hangú férfi és női beszélők is dominánsabbnak hatnak, mint a magasabb hangon beszélők. A mélyebb hang akár magasabb társadalmi pozícióhoz is vezethet. Más kutatások a mélyebb férfihang és a magasabb női hang vonzóbb benyomást keltő szerepét igazolták a párválasztásban (Borkowska–Pawlowski 2011; Puts et al. 2016).

A kutatók azt is vizsgálták, hogy a hallgatókban milyen – pozitív vagy negatív – benyomás, attitűd alakul ki az elképzelt beszélővel kapcsolatban, illetve a hallott hang által keltett benyomás alapján adott helyzetekben milyen döntéseket hoznak vagy hoznának. Ennek a kérdéskörnek jellegzetes alkalmazását jelenti a legmegfelelőbb reklámhangok kiválasztása. Chattopadhyay és munkatársai (2003) azt találták, hogy a gyorsabb tempóval, mélyebb hangon lejátszott reklámszöveg esetében a reklámozott márkát jobban kedvelték a hallgatók, azonban normál tempóval történt lejátszás esetén lényegesen jobban figyeltek a reklámra, és jobban is emlékeztek a tartalmára. Martín-Santana és kollégái (2016) kutatása szintén kimutatta, hogy a mélyebb hangú bemondó kedvezőbb hatást vált ki a hallgatókban, ráadásul a mély hangú férfi bemondók szignifikánsan nagyobb professzionalizmust és bizalmat keltettek, mint a többi bemondó. Konkrét döntéshelyzetet pedig Shang és Lu (2022) kutatása vizsgált. A kísérleti személyek magánhangzókat hallottak, és a magánhangzók kellemes vagy kellemetlen hangzása alapján el kellett dönteniük, hogy befektetnek-e fél jent vagy nem. A hanginger csak egy röviden ejtett magánhangzó volt, így a nyelvi üzenet és a hangsúlyozás nem befolyásolhatta a válaszokat. Az eredmények azt mutatták, hogy a vonzónak ítélt hang esetében a kísérleti személyek szignifikánsan nagyobb arányban döntöttek a befektetés mellett, mint amikor kevésbé vonzónak tartották a hangot.

Az említett példákban a kutatók természetes, manipulálatlan beszédmintákat használtak. A gépi beszéd és a manipulációk lehetőségének elterjedésével azonban felmerült a kérdés, hogy a hallgatókban kialakuló benyomások, a hang által megjelenített virtuális személlyel kapcsolatos attitűdök mesterséges, illetve manipulált beszéd esetében miként alakulnak. Németh Géza (2006) felvetette, hogy nem közömbös a szintetizált beszédet alkalmazó automata szövegolvasók – például smsolvasók – hangjának minősége, ugyanis ha egy telefonos szolgáltató kellemetlen hangú felolvasót használ, ez a szolgáltató vállalat egészének megítélését is befolyásolhatja. Számos kutató vizsgálta azt, hogy a felhasználók különböző alkalmazások használatakor miként viszonyulnak a virtuális személyekhez, különösen azok hangjához. Dickerson és munkatársai (2006) kutatásában az orvostanhallgatók virtuális pácienseket kérdeztek ki a panaszaikról. A hallgatók egy része természetes módon rögzített, más részük gépi beszédet használó virtuális beteggel kommunikált. A szerzők nem találtak különbséget abban, hogy a hallgatók milyen kérdéseket tettek fel, tehát a gépi beszéd is alkalmas volt az oktatási célok elérésére. Ahhoz azonban, hogy megtanulják, hogyan tegyék fel a megfelelő kérdéseket, úgy találták, hogy a gépi beszédben növelni kell az érzelmi kifejezőerőt.

Egy másik esetben szintén természetes és gépi beszédet értékeltek kísérleti személyek. A kutatók (Cabral et al. 2017) egy virtuális karaktert hoztak létre többféle változatban. Az eredmények azt mutatták, hogy a kísérleti személyek jobban szerették a természetes hangon megszólaló figurát, kifejezőbbnek tartották a hangját, továbbá érthetőbbnek ítélték a beszédét. Más tulajdonságok esetében (vonzóság, hitelesség, „emberszerűség") a természetes és a gépi beszéd nem váltott ki eltérő benyomásokat. Abdulrahman és Richards (2022) kutatásában viszont azonos mértékben szerették, tartották barátságosnak, kedvesnek, kellemesnek a természetes, illetve szintetizált hangon beszélő figurákat. Az érthetőség szempontjából azonban a természetes beszéd itt is magasabb pontszámot ért el.

2. KÍSÉRLETEK SZINTETIZÁLT ÉS MANIPULÁLT BESZÉDDEL

Az előzőek alapján felmerül a kérdés, hogy a beszéd hangzásában melyek azok a sajátosságok – például a tempója, az alaphang magassága –, amelyek leginkább felelősek a benyomáselemek kialakulásáért. Ezekre a kérdésekre természetes, illetve szintetizált és manipulált beszédminták segítségével is kereshető válasz. A beszédminták lejátszásával, a kísérleti személyek válaszainak statisztikai feldolgozásával feltárható, hogy mely akusztikai sajátosságok milyen benyomásítéletekkel járnak együtt. A szintetizált beszéd előnye az, hogy a kutató teljes egészében kontrollálhatja a beszéd hangzását, és elméleti megfontolások alapján precízen beállíthatja az akusztikai paramétereket. A szintetizált hang hátránya ugyanakkor, hogy a szintetizált hangokkal végzett kísérletek eredményei kevésbé általánosíthatók, mivel azok jóval ritkábban fordulnak elő a hétköznapi életben, mint a természetes beszéd.

Mullenix és munkatársai (2003) kutatásában természetes és szintetizált férfi és női beszédrészleteket hallgattak a hallgatók. A kutatók az elképzelt beszélővel kapcsolatos attitűdöket vizsgálták. A természetes női beszéd kedvezőbb benyomást keltett a hallgatókban, mint a szintetizált. A férfi és a női szintetizált beszéddel szembeni attitűdök összehasonlításakor az derült ki, hogy a hallgatókra a férfi beszéd tett kedvezőbb benyomást. A szerzők a természetes beszéddel nem végeztek számításokat, de a közölt nyers adatok azt mutatják, hogy a női beszélőt a hallgatók tájékozottabbnak, őszintébbnek, erősebbnek, meggyőzőbbnek vélték hangja alapján. Trouvain és munkatársai (2006) egy 31 szótagból álló tesztmondatot generáltak különböző változatokban egy beszédszintetizátor segítségével, és azt vizsgálták, hogy az alaphang magassága, tartománya, a bemondás tempója és hangereje hogyan befolyásolja a kísérleti személyekben kialakuló benyomást az elképzelt virtuális beszélő személyiségvonásairól, azaz mi befolyásolja azt, hogy őszintének, izgatottnak, kompetensnek, kifinomultnak, nyersnek gondolják-e a beszélőt.

Manipulált beszédminták használatakor a kutató természetes bemondást vesz alapul, amelyen egyetlen vagy néhány kiválasztott paramétereket változtat, miközben a beszéd összes többi paramétere változatlan marad. Manipulálható például úgy egy beszédminta, hogy az eredeti bemondás mellett létrehozunk kismértékben felgyorsított, illetve lelassított változatokat is. A gyorsítás és a lassítás megvalósítható úgy, hogy a hangszínezet, hangmagasság és más, az akusztikai élményért felelős paraméterek nem változnak, a beszéd hangzása természetes marad. Ilyen manipulációt alkalmaztak Skoog Waller és munkatársai (2015) az életkorbecslési kísérletükben. A szerző természetes bemondásokat felgyorsított és lelassított 10-10%-kal, majd azok lejátszása után megállapította, hogy a lassított beszéd idősebb, a gyorsított beszéd fiatalabb beszélő benyomását kelti. Groen és munkatársai (2008) pedig azt kutatták, hogy a beszélő nemének megítéléséhez a hallgatók milyen akusztikai kulcsokat használnak. A kutatók férfi és női ejtésben rögzített szavak alaphangját, illetve a magánhangzók hangszínezetét (formánsszerkezetét) manipulálták. A különféle alaphang–formánsszerkezet kombinációk lejátszásával megállapították, hogy mely hangzások teszik férfiassá, illetve nőiessé a beszéd hangzását a hallgatók számára.

Összefoglalva azt mondhatjuk, hogy a beszéd a verbális üzenet közvetítése mellett olyan akusztikai inger is, amelynek észlelésekor a hallgató benyomást alakít ki a beszélőről, azaz naiv módon elképzel, jellemez egy (virtuális) személyt. Ez a jellemzés együtt jár attitűdök kialakulásával, és döntéseket is megalapozhat, éppen ezért indokolt annak kutatása, hogy milyen akusztikai paraméterek milyen benyomásokért és attitűdökért felelősek.

Az itt bemutatott kutatásunkban is hasonló kérdésekre keressük a választ. Kísérleti személyeknek különböző tempójú és alaphangú férfi és női ejtésben elhangzó mondatokat játszottunk le, majd attitűdjeiket mértük fel.

A kutatási kérdéseink a következők voltak:

1. Van-e különbség a férfi és a női beszélővel kapcsolatban kialakult attitűdök között?
2. Befolyásolja-e a beszélő alaphangja és artikulációs tempója az attitűdöket?

A kérdésekre kapott válaszok segítségével pontosabban megérthetjük, hogy a hallgatókban a beszélő hangja alapján hogyan alakulnak ki egyes attitűdök. Ennek ismeretében pedig virtuális személyek, például asszisztensek kialakításakor céltudatosan lehet természetes hangbemondásokat kiválasztani, illetve azok hangzását átalakítani.

3. ANYAG ÉS MÓDSZER

3.1. Hangingerek

A kutatáshoz a BEA spontánbeszéd-adatbázisból választottuk ki a beszédmintákat (Gósy et al. 2012). A több lehetséges beszélő közül azt a férfit és nőt választottuk, akik a felvétel idején 22 éves egyetemi hallgatók voltak, illetve alaphangjuk közel állt az életkoruknak megfelelő szakirodalmi átlaghoz. Ennek meghatározásához kiválasztottunk egy-egy kb. egyperces, semleges érzelmi állapotban elmondott spontánbeszéd-részletet az interjúból, és a Praat program (Boersma–Weenik 2019) segítségével megállapítottuk az alaphangértékeket. A férfi esetében 113 Hz-et, a női beszélőnél pedig 193 Hz-et mértünk. Ezeket az adatokat összehasonlítottuk Gráczi és munkatársai (2020) eredményeivel – amelyek szerint a 20–29 éves korosztályban a férfiak alaphangja átlagosan 112,79 Hz (SD = 20,37 Hz), a nőké pedig 190,82 Hz (SD = 27,11 Hz) –, így megállapítottuk, hogy a kiválasztott beszélők alaphangja közel áll a szakirodalomban közölt átlagos értékhez, azaz nem beszéltek kirívóan mély vagy magas hangon. Mindezek mellett mindketten a sztenderdnek tekinthető budapesti köznyelvi nyelvváltozatot beszélték, beszédükben semmilyen más dialektusra jellemző sajátosságok nem voltak megfigyelhetők. Egyikük sem dohányzott; továbbá beszédprodukciós zavar, hangképzési probléma, beszédhiba nélkül beszéltek. A kutatáshoz a BEA felvételi protokoll szerinti „mondatismétlés" című feladatból használtunk 9-9 mondatot az egyes beszélők bemondásában. A feladat során az interjúkészítő 8-12 szóból álló mondatokat olvasott fel, amelyeket az interjúalanyoknak meg kellett ismételniük. A feladat magas szintű kognitív működést nem igényelt, elsősorban a beszélők rövid távú memóriáját vette igénybe. A mondatokat a beszélők semleges érzelmi állapotban ismételték meg, így az érzelmi állapotuk sem kelthetett kedvezőbb vagy kedvezőtlenebb benyomást.

A kiválasztott mondatokat külön-külön hangfájlokként elmentettük, a Praat program Manipulation funkciójának felhasználásával az alábbi módosításokat hajtottunk végre nyolc bemondáson:

- egy hangfájlon 4 félhanggal megemeltük, egy másikon félhanggal mélyítettük az alaphangot, azaz a természetesnél magasabb, illetve mélyebb hangzású változatokat hoztunk létre;
- egy hangfájl időtartamát 10%-kal megnöveltük, egy másikét 10%-kal csökkentettük, azaz a természetesnél lassabb, illetve gyorsabb változatokat hoztunk létre;
- az említett két manipulációt kombináltan is alkalmaztuk, így gyorsított-magas, gyorsított-mély, lassított-magas és lassított-mély bemondásokat is generáltunk.

E manipulációk csak a kiválasztott akusztikai paramétereket változtatják meg, más paraméterek változatlanul maradnak, így a beszéd hangzása nem torzul. Miután elkészültek a manipulált bemondások, minden hangfájl elején elhelyeztünk egy 0,5 másodperces, 440 Hz-es szinuszhullámot (figyelemfelhívó hangjelzést), majd egy 1 másodperces szünetet. Végül létrehoztunk egy állandó, de véletlenszerű lejátszási sorrendet, amelyben a férfi és a női bemondások rendszertelenül váltakoztak.

3.2. Kísérleti személyek, adatgyűjtés

A kutatásban 19–42 éves férfi és női egyetemi hallgatók vettek részt (medián = 22 év, 51 fő, 17 férfi, 34 nő). Mindannyian magyar anyanyelvűek, ép hallásúak voltak. A kísérlet a PTE Művészeti Kar tantermeiben multimédiás számítógépről professzionális hangszórókon lejátszva, csoportos lehallgatással történt, csendes körülmények között, zavaró tényezők nélkül. A kísérlet vezetője ismertette a kísérlet menetét, majd lejátszott három beszédmintát. Ezzel a hallgatók megismerkedhettek a feladattal, nem csupán az első lejátszáskor találkoztak azzal. Egyúttal a kutatás vezetője ellenőrizte, hogy minden hallgató tisztán hallja-e a lejátszott beszédmintákat. A beszédminták lejátszását a kutató irányította, és csak akkor indította el a következő hangfájl lejátszását, amikor az előző beszédmintával kapcsolatban minden válaszadó minden kérdésre válaszolt.

A kísérleti személyek hétfokú skálákon jelölték, mennyire tartják kellemesnek a hallott hangot, továbbá négy helyzetben attitűdjük viselkedéses dimenzióját mértük fel (szívesen beszélgetne-e vele, fordulna-e hozzá segítségért, vásárolna-e tőle a boltban, választáson szavazna-e rá). A skálákon az 1 a legkedvezőtlenebb, a 7 a legkedvezőbb választ jelentette.

3.3. Eredmények

Az adatok feldolgozása a következőképpen történt. Ebben a kutatásban csak nem paraméteres számításokat végeztünk. A Friedmann-tesztet alkalmaztuk minden egyes vizsgált tulajdonságra. A teszt szignifikáns eredménye azt mutatja, hogy valahol, valamely beszélők kedveltsége között szignifikáns eltérés volt. Ezt nem kerestük meg, hanem a Friedmann-teszt eredményeként előálló kedveltségi sorrenddel dolgoztunk tovább. A sorrendekben Mann–Whitney-féle U-próbával megállapítottuk, hogy a férfi és a női beszédváltozatok kedveltsége között van-e eltérés. További számítást végeztünk annak megállapítására, hogy a férfi és a női hang azonos manipulációi azonos módon befolyásolják-e az adott hang kedveltségét. Külön-külön felállítottuk a férfi és a női bemodások sorrendjét, majd a

Spearman-féle rangkorrelációs együtthatóval megvizsgáltuk, hogy a férfi, illetve női mondatokon végzett azonos manipulációk azonos pozícióba juttatták-e a bemondásokat a két, kedveltség szerinti sorrendben. A számítások részleteit a függelék tartalmazza, itt csak a végkövetkeztetéseket közöljük.

Először a hang kellemessége alapján keltett benyomásokat vizsgáltuk. A sorrendet a függelékben az 1. ábra szemlélteti. A legkellemesebbnek a gyorsított-mély női bemondást, a legkevésbé kellemesnek lassított-emelt férfi bemondást találták a hallgatók. Összességében a normálhoz képest mélyebb és gyorsabb beszédminták kellemesebb benyomást keltettek. A Mann–Whitney-próba azt igazolta, hogy a női beszélő hangját kellemesebbnek ítélték a hallgatók. A férfi és a női bemondások rangsorát külön-külön is összehasonlítottuk, azaz megvizsgáltuk, hogy a kellemesség szerinti sorrendben az azonos manipulációk azonos helyeken szerepelnek-e. A számítás nem igazolt szignifikáns kapcsolatot. Az adatok alaposabb elemzése során azonban kiderült, hogy az összefüggést a normál alaphangú lassított, illetve az emelt alaphangú gyorsított bemondások „zavarták meg". Az előbbi esetében a férfi, az utóbbinál a női bemondás kedveltsége négy, illetve öt pozícióval előbbre volt a rangsorban, mint az ellenkező nemű beszélő ugyanazon manipulációval módosított bemondása. Ettől eltekintve a mélyített-gyorsított beszédminta kedveltsége, illetve az emelt és lassított változatok elutasítása mindkét nem esetében megmutatkozott.

Hasonló eljárással megvizsgáltuk, hogy mennyire szívesen beszélgetnének a hallgatók az adott beszélővel, a hangja alapján. A 2. ábra mutatja a sorrend alakulását. A legszívesebben a gyorsított és normál alaphangú női bemondás beszélőjével, a legkevésbé pedig a lassított és emelt alaphangú férfi bemondás beszélőjével beszélgettek volna a hallgatók. A férfi és a női bemondások között nem szignifikáns, de tendenciaszerű különbség adódott. A férfi beszélő gyorsított beszédének normál és mélyített változata is kedvezőbb attitűdöt váltott ki, mint az előző esetben, ezek a női bemondások öt változatánál is magasabb helyre kerültek az összesített rangsorban. A női beszélő mélyített-lassított hangja ugyanakkora legkevésbé kedvelt hat bemondás közé került. A külön-külön felállított sorrendek összehasonlítása azt mutatta ki, hogy a gyorsítás és a mélyítés kedvezőbb, a lassítás és a hangmagasság emelése pedig kedvezőtlenebb attitűd kialakulását eredményezte a férfi és a női bemondásoknál is.

Harmadik lépésként a segítségkéréssel kapcsolatos attitűdöket mértük fel. A kapott sorrendet a 3. ábra mutatja. A hallgatók a legszívesebben a gyorsított és mélyebb alaphangú női bemondás beszélőjétől, a legkevésbé pedig a lassított és emelt alaphangú férfi bemondás beszélőjétől kértek volna segítséget. A sorrendből az is megállapítható, hogy a női hang alapján általában kedvezőbb, a férfihang alapján pedig általában kedvezőtlenebb attitűd alakult ki a hallgatókban az elképzelt beszélő iránt. A különbség szignifikáns volt. A férfi és a női felvételeket külön-külön is rangsoroltuk az attitűd alapján. E két rangsor összehasonlítása azt

mutatta, hogy ugyanaz a manipuláció hasonló módon befolyásolta az attitűdök alakulását mindkét nem esetében. A mélyített és a normál alaphangú, gyorsított bemondások iránti kedvezőbb attitűd itt is megmutatkozott, míg az emelt-lassított kombináció kedvezőtlenebb benyomást keltett mindkét nem esetében.

A következő számítások arra vonatkoztak, hogy a hallgató az elképzelt beszélőtől szívesen vásárolna-e boltban. A 4. ábra illusztrálja a beszédminták attitűd szerinti sorrendjét. A legszívesebben a gyorsított és normál alaphangú női bemondás beszélőjétől, a legkevésbé pedig a normál tempójú és emelt alaphangú férfi bemondás beszélőjétől vásároltak volna a hallgatók. A gyorsított és mélyített hangok iránti általános preferencia itt is megmutatkozott. A férfi és a női felvételeket külön-külön is rangsoroltuk az attitűd alapján. E két rangsor összehasonlítása azt mutatta, hogy ugyanaz a manipuláció hasonló módon befolyásolta az attitűdök alakulását mindkét nem esetében. A férfi és a női beszélők külön-külön megállapított sorrendjei azonban csak megközelítőleg mutattak szignifikáns összefüggést. Az adatok további elemzése során kiderült, hogy néhány esetben háromhelyezésnyi különbség is volt az egyes beszédminták kedveltsége között a két rangsorban. Az emelt-gyorsított és a manipulálatlan hang a női változatban kedveltebb volt, a mélyített-lassított viszont a férfi beszélő esetén volt kedveltebb.

Végül megvizsgáltuk, hogy a hallott hang alapján a hallgatók mennyire szívesen szavaznának a beszélőre egy választás alkalmával. A kapott sorrendet az 5. ábra illusztrálja. A legszívesebben a gyorsított és mélyebb alaphangú női bemondás beszélőjére, a legkevésbé pedig a normál tempójú és emelt alaphangú férfi bemondás beszélőjére szavaztak volna a hallgatók. Megfigyelhető, hogy a legkedveltebb hat beszédminta közül itt már három is a férfi beszélő valamely bemondása volt, ugyanakkor a női változatban létrehozott emelt-lassított beszédminta a legkevésbé kedveltek közé került. A Mann–Whitney-próba emiatt nem igazolt szignifikáns eltérést a férfi és a női bemondások kedveltsége között. Megvizsgáltuk, hogy a férfi és a női bemondások rangsorában az azonos manipulációk azonos helyeket foglaltak-e el. A két rangsor szignifikáns összefüggése arra engedett következtetni, hogy ebben az esetben is a mélyebb és gyorsabb bemondások kedvezőbb benyomásokat keltettek, míg a normálnál magasabb hangú és lassított hangok kevésbé voltak vonzóak a hallgatók számára mindkét nem esetében.

Végül arra a kérdésre kerestük a választ, hogy a kellemességbenyomás szerinti sorrend összefügg-e a különböző attitűdök szerinti sorrendekkel. A Spearman-féle korrelációs együtthatók minden esetben erős összefüggést mutattak (beszélne vele: $\varrho = 0{,}889$, segítséget kérne tőle: $\varrho = 0{,}928$, vásárolna tőle: $\varrho = 0{,}948$, szavazna rá: $\varrho = 0{,}862$, minden esetben $p < 0{,}001$). Ezek az adatok a holdudvarhatást igazolják, azaz a kellemesebb benyomást keltő hang együtt járt minden esetben a kedvezőbb döntéssel.

4. KÖVETKEZTETÉSEK

A kutatásunk kiindulópontja az volt, hogy az elhangzó beszéd nem csak verbális üzenetet közvetít, hanem a hang minőségének észlelésekor a hallgató elképzel egy virtuális személyt, akit a sztereotípiái alapján különféle tulajdonságokkal ruház fel, illetve különböző attitűdöket is kialakít vele kapcsolatban. Első kutatási kérdésünknek megfelelően a férfi és a női beszélő iránti attitűdök különbségét vizsgáltuk. A női beszélő bemondásai szignifikánsan kedvezőbb értékelést kaptak a kellemesség, a beszélőtől való segítségkéréssel, illetve vásárlással kapcsolatos attitűdök esetében. A társalgás iránti nyitottságnál közel szignifikáns volt a különbség a női beszélő javára, míg a szavazás esetében – bár az adatok egyszerű szemrevételezésével a női beszélő kedveltsége itt is feltételezhető volt – a számítás szignifikáns különbséget nem igazolt.

Második kutatási kérdésünk az alaphang és a tempó változtatásának szerepére irányult. Általános tendenciaként – megengedve némi szóródást – azt állapíthatjuk meg, hogy az átlagos alaphangú és tempójú bemondásokhoz viszonyítva a manipulált beszédminták közül a gyorsított-mély változatok általában kedvezőbb attitűdöt keltettek, mint a lassított-emeltek.

Eredményeink összhangban vannak Chattopadhyay és munkatársai (2003) közlésével. Ők azt találták, hogy a gyorsabban és mélyebb alaphangon elhangzott reklámszövegek esetében a hallgatókban kedvezőbb attitűd alakul ki a reklámozott márkával szemben, mint amikor a gyorsított-emelt bemondást hallották. Kísérletükben a normál tempójú, de magas vagy mély alaphangú bemondások nem mutattak különbséget. Lassított beszédet viszont nem használtak. A női hang iránti preferencia egybecseng Machado és munkatársai (2012) eredményével, azonban ők a magas hangú női beszélő kedveltségét mutatták ki. Igaz, ők szintetizált hangokat használtak, illetve a kontextus is más volt, hiszen a kísérleti személyeknek egy figyelmeztető bemondás *(Attention to the warning!)* alapján kellett a beszélőt jellemezniük.

Kutatásunkban több korlátozó tényezőről is említést kell tenni. Itt kizárólag nem paraméteres statisztikát alkalmaztunk, így például a független változók esetleges interakcióinak vizsgálatát sem végeztük el. További lényeges szempont, hogy bár a női hangot minden esetben szignifikáns vagy közel szignifikáns mértékben preferálták a hallgatók, ebből nem következik az, hogy a női beszédet általában is kedveltebbnek vélik a hallgatók. Az itt bemutatott kutatási eredmények kizárólag a kutatásban használt beszélőkre vonatkoznak. Nem foglalkoztunk továbbá az egyes beszélők egyedi hangszínezetével. Nem zárható ki, hogy a magánhangzók formánsszerkezete, az 1 kHz fölötti energia szisztematikusan befolyásolhatta még a döntéseket, és a zönge szabályosságát leíró paramétereket (jitter, shimmer), a jel/zaj arányt, de az irregulárisan képzett, „recsegő" zönge esetleges jelenlétét sem vizsgáltuk. Azt sem vettük figyelembe, hogy a női és a férfi hallgatók benyomá-

sai esetleg eltérnek-e. Wiener és Chartrand (2014) kutatása e két utóbb említett tényező jelentőségét jól példázza. Reklámszöveg lejátszásakor a férfi hallgatók számára nem volt jelentősége annak, hogy recsegő *(creaky)* zöngével beszél-e a reklámbemondó vagy sem, a reklámozott termék iránti érdeklődésük nem tért el szignifikánsan. A női hallgatók esetében azonban szignifikánsan magasabb volt a termék iránti érdeklődésük, ha recsegő zöngével beszélő férfi volt a bemondó, míg a legalacsonyabb a recsegő zöngével beszélő női bemondó esetében volt. Ilyen számításokkal a jelen kutatásunk is kiegészíthető, így árnyaltabb képet kaphatunk a vizsgált jelenségekről.

Az itt prezentált kutatásunk megerősítette azokat a korábbi eredményeket, hogy különböző akusztikai minőségű beszédminták különböző benyomásokat és attitűdöket alakítanak ki a hallgatókban, amelyek holdudvarhatás-szerűen együtt járnak. Ez a kutatás csak a kellemes benyomást és néhány hétköznapi életben előforduló döntéshelyzetben mérte fel a válaszadók attitűdjeit, de ezek összefüggése igazolódott. Indokoltnak látjuk a jelenség további vizsgálatát minden olyan helyzetre vonatkozóan, amikor egy szolgáltatás üzemeltetésekor, egy reklámban vagy akár a virtuális térben bármilyen körülmények között beszéd hangzik el. A kutatási eredmények ismeretében az elhangzott beszéd hallgatóinak attitűdje és magatartása megjósolható, illetve meghatározható, kiválasztható az adott helyzetben a legmegfelelőbb hangzású beszéd, céltudatosan tervezhetők virtuális karakterek.

SZAKIRODALOM

Abdulrahman, Amal – Richards, Deborah 2022: Is Natural Necessary? Human Voice versus Synthetic Voice for Intelligent Virtual Agents. *Multimodal Technologies and Interaction*, 6/7 (51): 1–17. https://doi.org/10.3390/mti6070051

Allport, Gordon W. – Cantril, Hadley 1934: Judging personality from voice. *Journal of Social Psychology: Political, Racial and Differential Psychology*, 5: 37–55.

Boersma, Paul – Weenik, David 2019: *Praat: Doing phonetics by computer*. www.praat.org

Borkowska, Barbara – Pawlowski, Boguslaw 2011: Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82: 55–59.

Cabral, João Paulo – Cowan, Benjamin R. – Zibrek, Katja – McDonnell, Rachel 2017: The Influence of Synthetic Voice on the Evaluation of a Virtual Character. In: *Interspeech 2017*. ISCA. 229–233. https://www.isca-speech.org/archive/interspeech_2017/cabral17_interspeech.html [2022. 10. 24.]

Chattopadhyay, Amitava – Dahl, Darren W. – Ritchie, Robin J. B. – Shanin, Kimary 2003: Hearing voices: The impact of announcer speech characteristics on consumer response to broadcast advertising. *Journal of Consumer Psychology*, 13/3: 198–204.

Dickerson, Robert – Johnsen, Kyle – Raij, Andrew – Lok, Benjamin – Stevens, Amy – Bernard, Thomas – Lind, D. Scott 2006: Virtual patients: assessment of synthesized versus recorded speech. *Studies in Health Technology and Informatics*, 119: 114–119.

Gao, Yang – Rita Singh – Bhiksha Raj 2018: Voice Impersonation using Generative Adversarial Networks. *arXiv*:1802.06840v1 https://arxiv.org/abs/1802.06840 [2022. 10. 18.]

Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Gráczi Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012: BEA: Beszélt nyelvi adatbázis. In: Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Budapest: Akadémiai Kiadó. 9–24.

Graber-Mitchell, Nicolas 2021: Artificial Illusions: Deepfakes as Speech (May 28, 2021). *Intersect* 14/3. https://ssrn.com/abstract=3876862 [2022. 10. 18.]

Gráczi Tekla Etelka – Gósy Mária – Krepsz Valéria – Markó Alexandra – Huszár Anna – Damásdi Nóra – Gocsál Ákos 2020: Az alapfrekvencia jellemzői az életkor és nem függvényében. In: Fóris Ágota – Bölcskei Andrea – Bóna Judit – Gráczi Tekla Etelka – Markó Alexandra (szerk.): *Nyelv, kultúra, identitás. Alkalmazott nyelvészeti kutatások a 21. századi információs térben: III. Fonetika*. Budapest: Akadémiai Kiadó. https://mersz.hu/hivatkozas/m675nyki3f_23. [2022. 09. 20.]

Groen, Wouter B. – van Orsouw, Linda – Zwiers, Marcel – Swinkels, Sophie – van der Gaag, Rutger – Buitelaar, Jan K. 2008: Gender in voice perception in autism. *Journal of Autism and Developmental Disorders*, 38: 1819–1826.

Herzog, Hertha 1933: Stimme und Persönlichkeit. *Zeitschrift für Psychologie* 130: 300–369.

Khanjani, Zahra – Gabrielle Watson – Vandana P. Janeja 2021: How Deep Are the Fakes? Focusing on Audio Deepfake: A Survey. *arXiv*:2111.14203 (2021). https://arxiv.org/abs/2111.14203 [2022. 10. 22.]

Machado, Sheron – Duarte, Emília – Teles, Júlia – Reis, Lara – Rebelo, Francisco 2012: Selection of a voice for a speech signal for personalized warnings: The effect of speaker's gender and voice pitch. *Work*, 41: 3592–3598. https://doi.org/10.3233/WOR-2012-0670-3592

Martín-Santana, Josefa – Muela-Molina, Clara – Reinares-Lara, Eva – Rodriguez-Guerra, Miriam 2015: Effectiveness of radio spokesperson's gender, vocal pitch and accent and the use of music in radio advertising. *Business Research Quarterly*, 18: 143–160.

Mullennix, John W. – Stern, Steven E. – Wilson, Stephen J. – Dyson, Corrie-lynn 2003: Social perception of male and female computer synthesized speech. *Computers in Human Behavior*, 19/4: 407–424. https://doi.org/10.1016/S0747-5632(02)00081-X

Németh Géza 2006: Az akusztikai arculat szerepe az infokommunikációs szolgáltatók megítélésében. *Híradástechnika*, LXI/8: 17–21.

Pear, Thomas H. 1931: *Voice and Personality*. London: Chapman and Hall. https://www.mpi.nl/publications/item2368438/voice-and-personality [2022. 09. 20.]

Puts, David A. – Hill, Alexander K. – Bailey, Drew H. – Walker, Robert S. – Rendall, Drew – Wheatley, John R. – Welling, Lisa L. M. – Dawood, Khytam – Cárdenas, Rodrigo – Burriss, Robert P. – Jablonski, Nina G. – Shriver, Mark D. – Weiss, Daniel – Lameira, Adriano R. –a Apicella, Coren L. – Owren, Michael J. –Barelli, Claudia – Glenn, Mary E. – Ramos-Fernandez, Gabriel 2016: Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society B* 283/20152830: 1–8. https://doi.org/10.1098/rspb.2015.2830

Sapir, Edward 1927/1991: Beszéd és személyiség. In: Szépe György (szerk.): *Az ember és a nyelv*. Budapest: Gondolat. 115–131.

Shang, Junchen – Liu, Zhihui 2022: Vocal attractiveness matters: Social preferences in cooperative behavior. *Frontiers in Psychology*, 13: 877530. https://doi.org/10.3389/fpsyg.2022.877530

Skoog Waller, Sara – Eriksson, Mårten – Sörqvist, Patrik 2015: Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology* 6/978). https://doi.org/10.3389/fpsyg.2015.00978

Stylianou, Yannis 2009: Voice Transformation: A survey. In: *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. Taipei, Taiwan: IEEE. 3585–3588. https://doi.org/10.1109/ICASSP.2009.4960401

Suire, Alexandre – Raymond, Michel – Barkat-Defradas, Melissa 2021: Voice, sexual selection, and reproductive success. In: Weiss, Benjamin – Trouvain, Jürgen – Barkat-Defradas, Melissa – Ohala, John J. (szerk.): *Voice Attractiveness*. Singapore: Springer Singapore. 125–138. https://doi.org/10.1007/978-981-15-6627-1_7

Trouvain, Jürgen – Schmidt, Sarah – Schröder, Marc – Schmitz, Michael – Barry, William J. 2006: Modelling personality features by changing prosody in synthetic speech. In *Speech prosody: 3rd international conference, Dresden, May 2–5, 2006*. Dresden. http://www.coli.uni-saarland.de/~trouvain/trouvain_etal2006.pdf [2022. 09. 20.]

Weirich, Melanie 2008: Vocal stereotypes. In: Botinis, Antonis (szerk.): *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics. 25–27 August, 2008. Athens, Greece.* Athens: University of Athens. 229–232. https://bit.ly/3RaijQJ [2022. 09. 20.]

Wiener, Hillary J. D. – Chartrand, Tanya L. 2014: The effect of voice quality on ad efficacy. *Psychology & Marketing*, 31/7: 509–517.

Zuckerman, Miron – Driver, Robert E. 1989: What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, 13/2: 67–82.

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal NKFIH-FK-128814 számú pályázata támogatta.

FÜGGELÉK

A függelék a kutatási eredmények részleteit tartalmazza. Az ábrákon szereplő két táblázat cellái a kétféle manipuláció szerint előállított bemondásokat jelentik. Vízszintesen a tempó, függőlegesen az alaphang manipulációja szerint rendeztük a cellákat. A középső cellák a manipulálatlan bemondásokat jelentik. A cellákban szereplő szám a rangsor szerinti sorszám, mégpedig az 1. számú a legkedveltebb, a 18. számú pedig a legkevésbé kedvelt beszédminta. A fehér háttér a legkedveltebb hat, a világosszürke háttér a középső hat, míg a sötétszürke háttér a legkevésbé kedvelt hat beszédmintát jelöli. Az 1. ábráról az olvasható le, hogy a legkellemesebb benyomást a mélyebb és gyorsított női hang, míg a legkevésbé kellemeset az emelt és lassított férfihang keltette.

18	17	16	magas	13	12	2
8	15	10	normál	7	5	3
14	9	6	mély	11	4	1
lassú	normál	gyors		lassú	normál	gyors
férfi				nő		

1. ábra. *A kellemességbenyomás szerinti rangsor.* A Friedman-teszt szignifikáns eredményt adott [$\chi^2(17) = 225{,}570$, $p < 0{,}001$]. A női beszélő hangja szignifikánsan kellemesebb benyomást keltett ($U = 13$, $p < 0{,}05$). A férfi és a női bemondásokat külön-külön is rangsoroltuk, és megvizsgáltuk, hogy az azonos manipulációk a kedveltség szerinti sorrendben azonos pozíciókban foglalnak-e helyet. A Spearman-korrelációs együttható ($\varrho = 0{,}577$, $p = 0{,}104$) nem igazolt szignifikáns kapcsolatot.

18	17	14	magas	9	11	3
7	15	6	normál	8	12	1
16	10	4	mély	13	5	2
lassú	normál	gyors		lassú	normál	gyors
férfi				nő		

2. ábra. *A beszélgetés iránti attitűd szerinti rangsor.* A Friedman-teszt szignifikáns eredményt adott [$\chi^2(17) = 178{,}774$, $p < 0{,}01$]. A Mann–Whitney-próbát elvégezve $U = 19{,}00$, $p = 0{,}058$ adódott, ami nem szignifikáns, de tendenciaszerű különbséget jelent a férfi és a női bemondások között. A Spearman-féle rangkorrelációs azt mutatta ki, hogy az azonos manipulációval előállított beszédminták a férfi és a női bemondások külön-külön megállapított sorrendje összefügg ($\varrho = 0{,}767$, $p < 0{,}05$).

18	17	14	magas	12	9	3
8	15	7	normál	11	5	2
16	13	4	mély	10	6	1
lassú	normál	gyors		lassú	normál	gyors
férfi				nő		

3. ábra. *A beszélőtől való segítségkérés iránti attitűd szerinti rangsor.* A Friedman-teszt szignifikáns eredményt adott [$\chi^2(17) = 200{,}815$, $p < 0{,}001$]. A Mann–Whitney-próba is megerősítette, hogy a női hang különböző változatai szignifikánsan kedvezőbb attitűdöt keltettek ($U = 14{,}00$, $p < 0{,}05$). A nemenként külön-külön felállított rangsorok összefüggést mutattak ($\varrho = 0{,}683$, $p < 0{,}05$).

17	18	14	magas	10	12	2
11	16	7	normál	8	4	1
15	9	6	mély	13	5	3
lassú	normál	gyors		lassú	normál	gyors
férfi				nő		

4. ábra. *A beszélőtől való vásárlás iránti attitűd szerinti rangsor.* A Friedman-teszt szignifikáns eredményt adott [$\chi^2(17) = 194{,}511$, $p < 0{,}001$]. A női beszélő szignifikánsan kedveltebb volt (Mann–Whitney $U = 13{,}00$, $p < 0{,}05$). A külön-külön felállított sorrendek tendenciaszerű, közel szignifikáns korrelációt mutattak ($\varrho = 0{,}650$, $p = 0{,}058$).

17	18	13	magas	14	11	8
10	15	5	normál	12	7	2
16	6	3	mély	9	4	1
lassú	normál	gyors		lassú	normál	gyors
	férfi				nő	

5. ábra. *A beszélőre való szavazás szerinti rangsor.* A Friedmann-teszt szignifikáns eredményt adott [$\chi^2(17) = 174{,}063$, $p < 0{,}001$]. A férfi és a női bemondások iránti attitűdök között nem adódott szignifikáns eltérés az összesített sorrendben ($U = 23$, $p = 0{,}122$). A külön-külön felállított sorrendek közötti rangkorrelációs együttható szignifikáns értéke ($\varrho = 0{,}783$, $p < 0{,}05$) arra enged következtetni, hogy a két sorrend összefügg.

GULD ÁDÁM

A deepfake és CGI-technológia az influenszermarketing szolgálatában: így formálják át a digitális karakterek az ismertségipar működését

A kortárs médiakörnyezetben egyre komolyabban kell vennünk a figyelmeztetést, hogy ne higgyünk a szemünknek. A 2010-es évek végétől digitálisan kreált influenszerek és deepfake-alkalmazások formálják át a hírnév és ismertség világát, mivel egyre nagyobb népszerűségnek örvendenek a digitálisan létrehozott vagy digitálisan manipulált influenszerek. A CGI- (computer-generated images, vagyis számítógép által létrehozott képek) és a deepfake (mesterséges intelligenciával támogatott, mélytanulás segítségével alkotott vizuális tartalmú) karakterek ma már nemcsak ártatlan játékszerek, hanem komoly kulturális és gazdasági hatással bíró megoldások is. A fejezetben azt mutatom be, hogy milyen technológiai megoldások működnek a digitálisan kreált karakterek mögött, ezeknek milyen típusait tudjuk megkülönböztetni, illetve hogyan reagál rájuk a közönség. Öt esettanulmányon keresztül értelmezem, hogy a megoldás milyen gazdasági potenciállal rendelkezik, valamint milyen fogadtatásra talál a fogyasztók körében.

Kulcsszavak: CGI-influenszerek, influenszermarketing, online híresség, hatás-befogadás

1. BEVEZETÉS

A médiában született hírnév sajátossága, hogy a technológia fejlődésével párhuzamosan az ismertség és a népszerűség viszonyai is folyamatosan változnak. Még csak néhány éve ismerkedünk az online híressé váló véleményvezérek jelenségével, ma pedig már ott tartunk, hogy digitálisan kreált deepfake- és CGI-karakterek válnak egyre népszerűbbé, elsősorban a Z generációba tartozó fiatalok körében (Guld 2021). Vajon még képesek vagyunk-e megkülönböztetni egy hírességről készült valódi képet vagy videót egy olyantól, amelyben egy CGI-influenszer szerepel? Vagy fel tudunk-e ismerni egy olyan hamisított karaktert, amely csak utánoz-

za egy ismert személy külsejét, beszédét, testbeszédét? Már az álhírek *(fake news)* és álprofilok is alaposan megkérdőjelezték az online média forrásainak megbízhatóságát, de a CGI-karakterek és a mesterséges intelligenciával manipulált videók terjedésével szinte már alig tudjuk eldönteni, kinek és minek hihetünk. Mire számíthatunk a jövőben? Milyen technológia áll az említett megoldások mögött? Mire alkalmasak ezek a lehetőségek? Milyen kockázatokat rejtenek magukban? Az alábbiakban ezekre a kérdésekre keressük a válaszokat.

A fotók és mozgóképek hamisítása önmagában nem újdonság, a vizuális tartalmak manipulációja szinte egyidős a fotózás és a film történetével. Az viszont jelentős változást jelent, hogy korábban egy ilyen tartalom előállításához, a felvett anyag megvágásához és hamisításához komoly technikai háttér, valamint jelentős szakmai tapasztalat kellett. Ehhez képest a CGI- és deepfake-alkalmazások, amelyek 2017-ben jelentek meg először a digitális piacon, szabadon hozzáférhető programok, amelyek néhány kattintás után lehetővé teszik a fotók és videók átszerkesztését, mindezt olyan minőségben, hogy a végeredmény egy hétköznapi felhasználó számára gyakorlatilag megkülönböztethetetlen a valóságtól. A deepfake-alkalmazások ma már bárki számára elérhetők, mobiltelefonon vagy asztali gépen is futtathatják őket a felhasználók. Mindössze néhány gigabájt tárhelyet igényelnek, az egyszerűbb alkalmazások használata pedig néhány perc alatt elsajátítható. 2018-tól egyre több ilyen applikációt dobtak piacra a fejlesztők, így jelent meg a FakeApp, a DeepFaceLab, a FaceSwap, a MyFakeApp vagy a kínai fejlesztésű Zao, amely már 2019 óta jelentős népszerűségnek örvend az internetezők körében (Whittaker et al. 2021).

A deepfake-alkalmazások felhasználása széles keretek között mozog, ezek között hasznos, ártalmatlan és kifejezetten veszélyes példákat is találunk. A fejlesztők egy részét eredetileg nemes cél vezérelte, ugyanis a deepfake-videókban akár néhány fénykép segítségével életre lehet kelteni ma már nem élő embereket. Az elképzelés szerint így Albert Einstein tarthatott volna interaktív fizikaórát középiskolásoknak, vagy Marilyn Monroe mesélhetett volna az ötvenes-hatvanas évek amerikai történelméről. Az utóbbi években azonban a mainstream szórakoztatóipar is szívesen nyúlt a lehetőséghez, így több televíziós műsorban jelentek meg humoros paródiák, amelyeket deepfake-alkalmazások segítségével állítottak elő. A deepfake-alkalmazásokban rejlő üzleti lehetőségek miatt a megoldás az ismertségiparon belül is gyorsan teret hódított, ennek bizonyítéka a deepfake- és CGI-influenszerek megjelenése. Ezenkívül több tucat olyan mobilapplikáció is létezik, amelyek vicces digitális játéknak foghatók fel, s amelyekkel a felhasználók barátaik vagy ismert emberek arca mögé bújhatnak.

Azonban a deepfake-alkalmazások térhódítása problémákat is eredményezett az utóbbi években, és a jelenség két szempontból rendkívül aktuális. Részben a technikai fejlődés most ért el egy olyan szintet, ami már bárki számára lehetővé teszi a hamisított videók létrehozását. Részben pedig egyre nagyobb teret hódít

a felhasználók körében a felületes olvasás és megtekintés, a kontextus ismerete nélküli, kritikát mellőző befogadás gyakorlata (Subramanian 2017). Így nemcsak egyre több hamisított tartalom kerül az online térbe, de egyre nagyobb az esély arra is, hogy ezek megtévesszék a közönséget (Chesney–Citron 2019).

A deepfake-technológiával történő visszaéléseknek már jelenleg is vaskos története van, ezek közül az egyik legtöbbször vitatott probléma, hogy az alkalmazás szélesebb elterjedésével együtt a világhálót elárasztották a deepfake segítségével manipulált pornográf tartalmak. Ennek jellemzően híres amerikai színésznők estek áldozatául, akiknek az arcvonásait pornósztárok arcára másolták, így azt az illúziót keltették, mintha a felnőttfilmekben például Scarlett Johansson vagy Angelina Jolie szerepelnének (W1). Nem kellett sokat várni arra sem, hogy az alkalmazás a politikai manipuláció eszközévé váljon (Dobber et al. 2020). 2019-ben az amerikai politikusról, Nancy Pelosiról jelent meg egy manipulált videó, amelyben úgy tűnik, mintha a politikus ittasan nyilatkozna egy talk show-ban (W2). Eközben Indiában egy politikai lejáratókampányban került elő egy olyan szexvideó, amellyel kapcsolatban felmerült a deepfake használatának a gyanúja. A videón egy parlamenti képviselő (vagy hozzá hasonló férfi) látható szexuális aktus közben egy másik férfi társaságában. A politikus később az egyik parlamenti ülés során sírva magyarázta, hogy tönkre akarják tenni az életét, mivel a videóban valójában nem ő szerepel (W3). Az eset végkimenetelével a világsajtó már nem foglalkozott, de az incidens arra egyértelműen felhívja a figyelmet, hogy ehhez hasonló esetek tömegére lehet számítani a jövőben. A legnagyobb probléma pedig az, hogy azok a személyek, akiknek a neve egyszer bepiszkolódik egy botrány kapcsán, mindig magukon fogják viselni a stigmát, függetlenül attól, hogy a vádak igazak voltak-e vagy sem.

2. DIGITÁLISAN KREÁLT INFLUENSZEREK: REALISZTIKUS, STILIZÁLT ÉS CELEBRITÁSKARAKTEREK

A deepfake-technológia megjelenésével együtt az ismertségiparban is komoly változásokra számítanak a szakemberek, s ez a digitálisan megalkotott influenszerek alkalmazásának elterjedésével hozható összefüggésbe (Guthrie 2020). A 2010-es évek elejétől jellemző az a trend, hogy az online felületeken népszerűvé váló tartalom-előállítók (influenszerek vagy hétköznapi híresség) egyre nagyobb hatással vannak az ismertségipar működésére. A különböző közösségimédia-felületekre készített tartalmak segítségével ismertté váló hétköznapi fiatalok növekvő befolyására gyorsan felfigyelt a kommunikációs és marketingszakma – ezt jól igazolja az influenszermarketing és az online híresség eket menedzselő ügynökségek felvirágzása. Az influenszerek többsége azonban még ma is csak tizenéves, legfeljebb fiatal felnőtt. Ők pedig gyakran megbízhatatlanok, sokszor nem vagy nem az elvárt

módon teljesítik a partnerek megbízásait, ami sok fejtörést okoz az együttműködő márkáknak (Klausz 2019). Részben ezzel a problémával összefüggésben jelentek meg az első digitálisan kreált influenszerek 2016-ban, majd az igazi robbanást ebben a témában 2018 és 2019 hozta el (Callahan 2021).

De kik is azok a digitális influenszerek? A digitálisan megalkotott karaktereknek alapvetően három típusát tudjuk megkülönböztetni. A valósághű, számítógép által előállított karakterek, angolul *realistic CGI* figurák legfontosabb jellemzője, hogy szinte teljes mértékben valósághűek. E karaktereket jellemzően kereskedelmi céllal, például divatmárkák megrendelésére hozzák létre, s ennek megfelelően olyan megbízható reklámfelületekként működnek, amelyek minden helyzetben a márka szándékainak megfelelően viselkednek. A valósághű CGI-szereplők részletesen kidolgozott fiktív személyes háttérrel és kapcsolatrendszerrel rendelkeznek, a valós influenszerekhez hasonlóan gyakran keverednek kalandokba: szerelmesek lesznek, összevesznek, kibékülnek egymással. Egyre gyakrabban alkalmazzák őket olyan influenszer-együttműködésekben, ahol a CGI-karakter egy valódi hírességgel együtt jelenik meg az online térben.

Ezzel szemben a stilizált CGI, angolul *stylized CGI*-karakterekkel nem azt a hatást szeretnék elérni, mintha azok valóságosak lennének. Éppen ellenkezőleg, ebbe a kategóriába olyan szereplőket sorolunk, amelyek leginkább rajzfilmfigurákra emlékeztetnek. Stilizált CGI-karaktereket, avatárokat ma már szinte bárki létrehozhat egy okostelefon segítségével. Ezeket többnyire nem is arra használják, hogy a valóságban nem létező, új karaktereket alkossanak meg, hanem valós személyeket helyettesítenek az online térben, például a Snapchat vagy a Facebook felületén. A stilizált CGI-avatárok marketingcélú hasznosítása egyelőre szűkös keretek között mozog, de a szakemberek nagy potenciált látnak a megoldásban.

Harmadikként említhetjük a celebritás-deepfake-et, amely létező híresség digitálisan manipulált képeit jelenti, ebben az esetben egy statiszta arcvonásait cserélik ki egy sztár képére. A technológia gyakorlati alkalmazására már évek óta láthatunk különböző példákat, ezek közül a legismertebb a Malaria No More elnevezésű jótékonysági szervezet 2019-es kampánya, amely David Beckham közreműködésével készült el (W4). Az eset jól mutatja, hogy amennyiben a deepfake-alkalmazást megfelelő keretek között használják, akkor az időt és pénzt takaríthat meg az alkotóknak, illetve oly módon lehet áthidalni nyelvi és kulturális különbségeket, amire korábban nem volt lehetőség. A Beckham közreműködésével készült maláriaellenes kampányfilmben például a sztár kilenc különböző nyelven szólt a nézőkhöz. Ehhez a hírességnek nem kellett kilenc nyelven megtanulnia a szöveget, hanem a deepfake-megoldás segítségével érték el azt a hatást, mintha valóban ő beszélne (W5).

Az aktuális marketingtrendeket vizsgálva azt találjuk, hogy a személyes vonzerőnek, a híres emberek meggyőző erejének egyre nagyobb szerepet tulajdonítunk: lényegében ez jelenti az influenszermarketing alapját (Appel et al. 2020).

A deepfake-alkalmazások viszont komoly kihívást és veszélyt is jelentenek ezen a területen. A deepfake az influenszermarketing új korszakát jelentheti, hiszen a jövőben olyan problémákkal és visszaélésekkel találkozhatunk, amelyek a márkákat és az influenszereket is negatívan befolyásolhatják. Példának okáért híres emberek, influenszerek beleegyezése nélkül készülhetnek olyan videók, amelyekben az arcukkal olyan termékeket népszerűsítenek, amelyekről valójában még nem is hallottak. Ez egyfelől az influenszer hitelességét rombolja, másfelől azoknak a márkáknak az imázsát is negatívan befolyásolhatja, amelyekkel a szereplőnek valóban van együttműködése. Az influenszerek személyére irányuló támadásra már most is akad példa. Az egyik legkorábbi eset még 2019-ből származik, amikor egy deepfake-videóban Kim Kardashian nyilatkozott arról, hogy mekkora örömmel manipulálja a közönségét azért, hogy még gazdagabb legyen. A videó pillanatok alatt elterjedt a világhálón, és nagyon sokan egy percig sem kételkedtek abban, hogy valódi-e (Davenport et al. 2020).

Az előzőekből is kiderül, hogy a deepfake- és CGI-karakterek jelentős része egyelőre megkülönböztethető a valós hírességektől. Ugyanakkor a deepfake-videók leleplezése a technológia fejlődése miatt már most is egyre komolyabb kihívást jelent (Whittaker et al. 2020). Egyfelől a fejlesztők egy csoportja versenyt fut az idővel, és olyan digitális alkalmazásokon dolgozik, amelyek képesek felismerni a hamisított videókat. Másfelől valószínűsíthető, hogy a technológia további fejlődésével a deepfake-videók még jobb minőségűek lesznek, ezért kiemelt esetekben, például politikai céllal készült hamisított videók leleplezésében, olyan szakemberekre is számítanak, mint a nyelvészek vagy a testbeszéd elemzésével foglalkozó kommunikációs szakemberek (Maras–Alexandrou 2019).

A legfontosabb feladat jelenleg mégis az, hogy belássuk, a CGI- és deepfake-technológia már itt van, és nem is fog eltűnni. Itt is igaznak tűnik a *megszokni vagy megszökni* tétel, bár megszökni egyre nehezebb, így jobb, ha megtanulunk együtt élni a technológiával. E megoldások komoly kihívásokat és lehetőségeket is rejtenek magukban a kommunikációs és marketingszakemberek számára (Vaccari–Chadwick 2020). Viszont lényeges, hogy ők még nagyobb figyelmet fordítsanak a márkák és a közönség edukálására, illetve folyamatosan tájékozódjanak az alkalmazással kapcsolatos jogi környezet változásairól. A hétköznapi felhasználók pedig ne higgyenek a szemüknek, legalábbis ne elsőre. A CGI- és deepfake-videók korában minden esetben érdemes megerősíteni a kritikai hozzáállást, és lehetőség szerint több forrásból tájékozódni a kétes eredetű tartalmak megbízhatóságát illetően (Kay et al. 2020).

A fejezet második felében öt rövid esettanulmányon keresztül mutatom be, hogy 2022-ben hol tart a CGI-influenszerek piaca. A példákon keresztül nem célom egy reprezentatív kép megrajzolása, pusztán azokat a lehetőségeket szeretném felvillantani, amelyek kulturális vagy piaci szempontból jól érzékeltetik a technológiában rejlő változatos lehetőségeket. Az itt vizsgált öt karakter alkotói

eltérő célokkal alkalmazták a CGI-influenszereket, ami a példákban közös, hogy az új digitális eszközök olyan megoldásokra adtak lehetőséget, amelyek hús-vér influenszerekkel aligha lettek volna működőképesek. A vizsgálatban bemutatott öt CGI-influenszer között a következő karaktereket találjuk: a divatmodell Ronald F. Blawko, a grafikus Sylvia Novack, a zenész Knox Frost, a bébiinfluenszerként pozicionált Elis baba és a KFC gyorsétterem figurája, Virtual Colonel.

A vizsgálat első lépéseként az influenszerek karakterére és csatornáira fókuszálok, miközben online források alapján igyekszem bemutatni a megoldások főbb jellemzőit. A második lépés alapvetően a kvalitatív tartalomelemzés módszerét követi, a kutatás pedig az influenszerek Instagram-felületén megosztott posztokra, képekre és szövegekre, valamint az azokhoz fűzött felhasználói kommentekre fókuszál (Mayring 2004; Sztompka 2009). A kommentek vizsgálatához alkalmazott módszer a kritikai diskurzuselemzés hagyományaira támaszkodik, az online vitafórumok sajátosságait figyelembe véve (Glózer 2007, 2013; Császi 2011). Bár a kutatás kvalitatív jellegű, a jellemző trendek, arányok és preferenciák kimutatása érdekében egyes összefüggéseket számszerű adatok is alátámasztanak. A jelentős elemszám miatt a vizsgálatban a csatornákra kihelyezett legjellegzetesebb, legnagyobb nézettséget vagy legnagyobb aktivitást generáló posztok jelennek meg. A következőkben a folyamatban lévő munka első eredményeit foglalom össze.

3. ÖT ESETTANULMÁNY

3.1. Z generációs szexszimbólum a maszk mögött: Ronald F. Blawko

Ronald F. Blawko az egyik első olyan CGI-influenszer, aki a divatiparban vált sikeressé, a karakter tevékenysége hosszú éveken keresztül befolyásolta az aktuális divat- és szépségtrendeket. Blawko megalkotói, Trevor McFedries és Sara DeCou, az amerikai Brud vállalat tulajdonosai nagyon korán felismerték a CGI-influenszerekben rejlő üzleti lehetőségeket, így elsők között léptek a piacra digitálisan kreált online véleményvezérekkel (W6). A cég Blawko mellett több híresebb karaktert is megalkotott, többek között a világ egyik legbefolyásosabb CGI-influenszereként számontartott Lil Miquelát vagy Bermudát, aki egy ideig Blawko barátnőjeként tűnt fel a közösségi oldalakon. A Brud cég termékeit, vagyis a virtuális influenszereket, a Huxley elnevezésű PR- és kommunikációs cég képviseli, amely az utóbbi években elérte, hogy a digitális karakterekkel a legnagyobb divatmárkák dolgozzanak együtt, sőt az együttműködésekhez számos hús-vér sztárt is sikerült megnyerni.

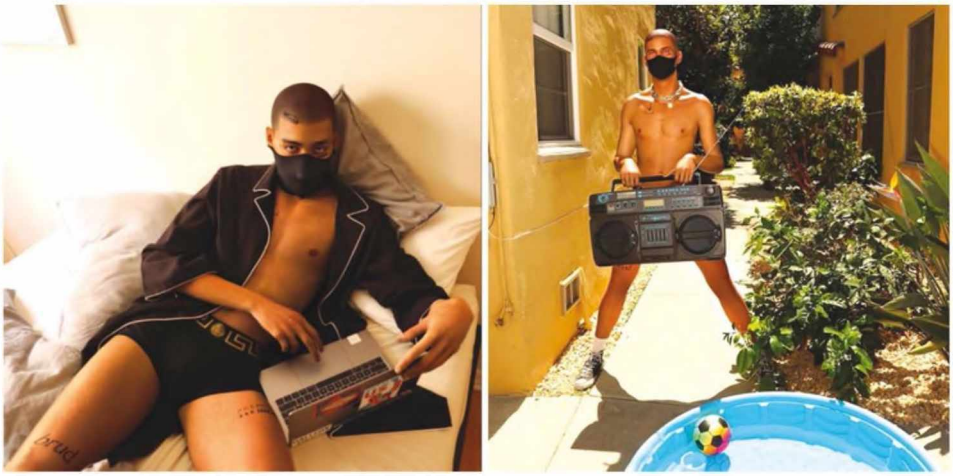
Ronald F. Blawko karaktere 2016-ban született, és azon kevés CGI-modellek egyike, aki az Instagramon, a Twitteren és a YouTube-on is ismertté vált @blawko22 néven (W7). Bár a karakter nem aktív, Instagramon még mindig 135 000

követővel rendelkezik, a legnépszerűbb posztjain közel 30 000 lájk található. Az alkotók szerint Blawko egy szintetikus robot, aki azonban több, mint egy vizuális konstrukció, mivel önálló és egyedi személyiséggel rendelkezik, van benne érzelem és „valódi" élet (W8). Blawko 2020 szeptemberéig kifejezetten aktív volt a közösségi médiában, korábban gyakran megjelent a YouTube-on, fotókat osztott meg magáról Instagramon, s rendszeresen részt vett az online közösségekben zajló eseményeken. Ezenkívül gyakran posztolt Twitteren, ahol a rajongói nyomon követhették, hogyan érzi magát, milyen gondolatok foglalkoztatják, éppen kikkel tölti az idejét (W9).

A karakter aktív médiajelenléte mögött természetesen a digitális marketinggel foglalkozó anyacég tudatos építkezése húzódik meg, Blawko megjelenése jól irányzott marketingeszköznek tekinthető. A siker titka az, hogy Blawko a kortárs populáris és ifjúsági kultúrákban divatosnak számító karakter, akinek a megjelenését és stílusát, személyiségét és viselkedését olyan elemekből gyúrták össze, amelyek külön-külön számos, Z generációs hírességben felfedezhetőek. A végeredmény egy olyan figura, ami más-más okokból kifolyólag a lányok és fiúk körében egyaránt népszerű, egy trendi férfitípusról van szó, amelyet a rajongói hozzászólások alapján az angol *fuckboy* (manipulatív, szívtipró nőcsábász) kifejezés ír le a legérzékletesebben.

Blawko vonzerejének legfontosabb tényezője a karakter megjelenése. A figura származására vonatkozóan csak találgatások vannak, de a színes bőrű férfi egzotikus kinézete és nyelvhasználata afrikai vagy latin-amerikai háttérre utalhat. Megjelenése alapján Blawko 18–20 év közötti fiatal felnőtt, aki nem túl izmos, de divatos, sportos fizikummal rendelkezik, a testét több helyen jellegzetes tetoválások díszítik. Az esztétikus férfitest bemutatása gyakran központi témája a posztoknak, a tartalmak egy része kifejezetten erotikus jellegű, ezekben a karakter sokszor erősen hiányos öltözékben látható, gyakran provokatív, szexuális aktust imitáló pózokban. Blawko frizurája, arcberendezése, arcának karaktere is tökéletesen illeszkedik az aktuális divatideálokhoz: géppel nyírt rövid haj, meredek szemöldökív és az arcán több helyen arctetoválások láthatók, a homlokán lévő tetoválás például Ashley Funicello Spinelli, a Walt Disney Stúdió animációs karaktere iránti szerelmét jelképezi (W10). Blawko feltűnő megjelenésének legmeghatározóbb kelléke egy maszk, ami az összes képen elfedi az arcának alsó harmadát. Ez a kiegészítő a felhasználói reakciók alapján különösen vonzóvá teszi a figurát, akinek esetében az esztétikus fizikummal párosított titokzatos megjelenés igazi szexszimbólumot eredményezett a Z generációs követők körében.

Blawko megjelenésének lényeges elemei a különböző ruhák és kiegészítők, amelyek szintén mindig a legutolsó divatot tükrözik. A karakter megjelenésére jellemző, hogy gyakran visel streetwear stílusú ruhákat, amelyeken általában egy-egy nagyobb márka logóját is láthatjuk. Blawko olyan márkáknak is „dolgozott" már, mint az Off-White, a Balenciaga és a Supreme. A tökéletes megjelenésre való



1. ábra. Két jellegzetes poszt Blawko Instagram-oldaláról (forrás: W7)

törekvés a posztok állandó témája, amelyekben Blawko narcisztikus hajlamai is egyértelműen felszínre kerülnek. A posztokból kiderül, hogy a karaktert lenyűgözi a saját kinézete, egy vallomásából megtudhatjuk, hogy legszívesebben klónozná magát, hogy csókot adhasson a saját homlokára. Más tartalmakból az derül ki, hogy a megjelenése egész nap élénken foglalkoztatja, többek között óriási hangsúlyt fektet a bőrápolásra, ezért a posztokban rendszeresen feltűntek a Glossier Skincare termékei (W11).

Blawko legtöbb követőt számláló közösségi oldala az Instagram, ahol az első poszt 2018. január 19-én jelent meg. Ebben a posztban is egy kevéssé burkolt termékelhelyezés látható, ugyanis a karaktert háttal láthatjuk, egy Supreme feliratos kabátban. Az Instagram-megjelenésekben folyamatos márkajelenlét figyelhető meg, szinte nincsen olyan poszt, amelyben ne tűnne fel valamelyik nagyobb világmárka terméke. A tartalmak rendszertelenül tűntek fel a felületen: voltak időszakok, amikor hetente többször vagy akár naponta kerültek ki tartalmak az oldalra, máskor egy vagy két hónap után került ki új poszt. A karakter legtermékenyebb éve 2019 volt, ekkor gyakran posztolt közös képeket Lil Miquelával, valamint az ex-barátnőjével, Bermudával. A karakterek kapcsolatára jellemző, hogy az influenszerek egymás Instagram-posztjait is lájkolták, s nemritkán kommentet is fűztek a megjelenésekhez, ahogy a követők is jelentős aktivitást mutattak. Ez nem véletlen, hiszen a posztokban gyakran megjelentek erősen provokatív, szexuális utalásokat tartalmazó megnyilvánulások.

Sok mai sztárhoz hasonlóan Blawko a YouTube-on is aktív volt, ott 4800 feliratkozó követte a videóit (W12). A videók jellegzetessége a karakter eltorzított hangja, amely robot lényéből fakad, és ami a felhasználói visszajelzések szerint először nagyon fülsértő az embernek. A beszéd másik jellegzetessége az erős af-

roamerikai akcentus, amelynek a megértése komoly kihívást jelenthet első hallásra. A csatornán posztolt videók tematikája változatos, ezek a beharangozó alapján egy fiatal robot életéről szólnak. A videókban Blawko szerelmi tippeket ad, vagy útmutatást nyújt ahhoz, hogyan lehet vonzó testet építeni. A nőcsábászként pozicionált karakter mindent tud a szerelemről, a videókban számos tipped ad arra, hogyan lehet a legegyszerűbben elcsábítani a nőket, vagy mi kell ahhoz, hogy az online randevúzás a Tinderen sikerrel járjon. A feltöltésekből az is kiderül, hogy Blawko is szeret randevúzni, és a környezetéből számos vonzó nővel volt kalandja, sőt egy bizarr szerelmi háromszög egyik tagjaként párhuzamos kapcsolata van Miquelával és Bermudával.

A követői aktivitást tekintve Blawko kifejezetten népszerű karakter, a rajongói között nagy számban vannak jelen nők, s a vizsgált kommentek többsége is (84%) nőktől származik. A hozzászólások többsége (71%) elismerő, rajongói hangnemben íródott, sokan kifejtik, hogy mennyire felnéznek Blawkóra, mennyire csodálják a stílusát, a testét, a megjelenését, a viselkedését. A kommentek megfogalmazásából világosan látszik, hogy a rajongók valós személyként gondolnak Blawkóra, s ha tudják is, hogy egy CGI-karakterről van szó, a játék kedvéért erről csak ritkán tesznek említést. A hozzászólásokban az érzelmi kötődés is megfigyelhető, ezekből kiderül, hogy a kommentelők egy része úgy tekint Blawkóra, mint egy idolra, egy tökéletes partnerre, akibe a rajongók bele is tudnak szeretni. Ezt tükrözik a következő hozzászólások: „Ha valódi ember lenne, simán belezúgnék"; „Szeretlek, gyönyörű vagy"; „Szeretlek, király vagy, annyira imádlak". A siker ellenére Blawko karaktere 2020 szeptembere óta nem aktív, mintha eltűnt volna a föld színéről, ennek okát és körülményeit az alkotók mindeddig nem tisztázták. 2023 februárjában az Instagram-oldala 135 000 követőt számlál, s mivel az utóbbi két évben az oldalon nem volt aktivitás, ez a szám az utolsó posztja óta folyamatosan csökken.

3.2. Egy CGI-influenszer élete és halála, avagy harc az időskori diszkrimináció ellen: Sylvia Novack

Sylvia Novack egy számítógép által generált karakter, ami a valósághű CGI-influenszerek kategóriájába tartozik, bár egyes megjelenései közelebb esnek a stilizált CGI-figurákhoz (W13). Sylvia egyike volt az első gyorsan öregedő CGI-karaktereknek, akinek a megszületésétől a haláláig mindössze öt hónap telt el, a karakter halálát 2020 novemberében jelentették be. Sylvia Instagram-profilja 2023. február 10-én összesen 2761 követővel rendelkezik, így bár soha nem tartozott a legismertebb CGI-influenszerek közé, szakmai körökben mégis jelentős visszhangot kapott a ténykedése. Sylvia a legtöbb CGI-karakterhez hasonlóan komplex, kidolgozott háttérrel és élettörténettel rendelkezett, grafikusként végzett, és éppúgy, mint egy valós influenszernek, voltak barátai, szerelmei és hobbijai is. Szeretett utazni, ked-

venc utazási célpontja Velence volt Olaszországban. Sylvia emellett gyakorlott jógázó volt, jógaoktatói oklevelet is szerzett, többször posztolt is ebben a témában tartalmakat. Személyiségét olyan hétköznapi apróságok is színesítették, mint hogy Sylvia a kávé megszállottja volt, az Instagram-bemutatkozásában erről ezt olvashatjuk: *Coffee operated robot, living her best life* (Kávéüzemű robot, aki igyekszik kihozni az életből a maximumot). A hagyományos online véleményvezérekhez hasonlóan ő is részt vett kereskedelmi együttműködésekben, amelyek során számos divatcéggel és nemzetközi hírű márkával dolgozott együtt. A vele kapcsolatban megjelent cikkek tanulsága szerint a virtuális divatvilág egyik úttörő múzsája volt, aki sok virtuális divattervező darabját mutathatta be rövidke élete során (W14).

Sylvia megalkotója a médiaművész Ziv Schneider, aki arra vállalkozott, hogy a karakter létrehozásával kritikát fogalmazzon meg az időskori diszkriminációval szemben (W15). Schneider érvelése szerint a választása azért esett az *ageism* problémájára, mert az időskori megkülönböztetés egyre jellemzőbb napjainkban, és a jelenség csak annyiban különbözik a társadalmi kirekesztés más formáitól, hogy társadalmilag általánosan elfogadott, és nem vagy csak nagyon ritkán lépnek fel ellene. Schneider a halál témáját is tabumentesíteni akarta a projekttel és rámutatni arra, hogy túlságosan hozzászoktunk ahhoz, hogy a halál és a haldoklás nem jelenik meg a vizuális kultúránkban. Schneider kifejtette, hogy mivel a halál természetes módon követi az öregedést, így gyakran mindkettőt kizárjuk a médiareprezentációból, ami természetellenes viszonyt teremt az élet egyik legtermészetesebb eseménye kapcsán (W16). Schneider szándékának megfelelően, Sylvia lett az első virtuális influenszer, aki elképesztő sebességgel idősödött az Instagramon, ahol a rajongói 2020 júliusától 2020 novemberéig követhették az életét egészen a haláláig. Sylvia a 30-as éveiben járó fiatal nőként jelent meg először a követői előtt, az első poszt 2020. július 4-én került fel az Instagramra (W17). A követők végig-

2. ábra. Sylvia egyik első posztja 30 évesen, valamint a halála előtti órákban 2020. november 20-án (forrás: W17)

nézhették, ahogy a virtuális karakter élete a szemük előtt bontakozott ki, miközben Sylvia öt hónap alatt öt évtizedet öregedett.

Sylvia az öt hónap alatt az Instagram-profiljára 81 posztot helyezett ki, amelyeken pontosan nyomon követhető az, ahogyan a karakter háromnaponta öregedett egy évet. A posztokban felváltva láthatunk képeket az idősödő nőről és az életmódjával kapcsolatos tipikus élethelyzetekről. A képeken láthatjuk őt jógázás vagy napozás közben, máskor biciklizik a szabadban vagy dolgozik, de olyan posztot is találunk, ahol Sylvia a Metropolitan Múzeumba látogat el. A történetek egy élet jellegzetes apró bosszúságait és örömeit is feldolgozzák, így azt is, amikor Sylvia összetűzésbe keveredett a szomszédaival, de az egyik barátjától kapott designerfülbevaló mégis feldobta a napját.

A megosztott tartalmakban az írásos szöveg legalább olyan hangsúlyos, mint a fotók, ezeknek a felét Schneider írta, s ezek olyan szövegek voltak, amelyek előrevitték a karakter történetét. Azonban minden második poszt egy számítógép által generált szöveg volt, ezeket a tartalmakat egy olyan számítógépes program alkotta meg, amellyel előzetesen befolyásos influenszerek profilját elemezték, így a rendszer megtanulta a legsikeresebb véleményvezérekre jellemző kommunikációs technikákat, stílusokat. Az algoritmust úgy alkották meg, hogy az általa írott szövegek is „öregedtek", az idősebb felhasználókra jellemző nyelvhasználatot tükrözték, vagyis egy elképzelt idősebb célcsoport igényét próbálták kielégíteni. Az idő előrehaladtával igyekeztek egyre bölcsebb tartalmakat gyártani, olyanokat, amelyek egy idősödő nőre lehetnek jellemzőek. Az elképzelés olyan jól sikerült, hogy sokan nem ismerték fel, hogy a szövegek egy részét számítógépes programok alkották. Schneider szakemberek segítségét is igénybe vette a projekt során, így az elöregedő társadalmak egészségi állapotát kutató tudósok segítettek az öregedés folyamatának és az életpálya alakulásának hiteles bemutatásában. A cél az volt, hogy Sylvia élete minél hitelesebben ábrázolja az öregedés egyes szakaszait és annak várható következményeit a kortárs társadalmi valóság kontextusában.

A projekt alapvető célja az időskori diszkrimináció, vagyis az *ageism* elleni küzdelem volt, s ennek egyik leglátványosabb mozzanatának azt tekinthetjük, amikor Sylvia Novack 2020 októberében interjút adott a *Cultured* magazinnak. A cikk *Tippek az öregedés ellen egy virtuális halandótól* címet kapta, amelyben Sylvia megfogalmazta azokat az elveket, amelyek később a *Dolgok, amiket az életről tanultam* című könyvében jelentek meg. Bár a történet szerint Sylvia a könyvét befejezte, halála előtt már nem tudta azt kiadni, viszont az a mai napig elérhető online. A korábban említett számítógépes program által írt szövegben a következő fejezetcímekkel találkozhat az olvasó: 1. Ne legyél egy wannabe; 2. Gondolj a gyerekekre; 3. Tudd meg: manapság az erős, szerető anya a menő; 4. Légy felelősségteljes; 5. Éld az életed; 6. Dolgozz keményen; 7. Élj a jelenben; 8. Szerezz barátokat; 9. Légy jelen; 10. Légy emberi lény; 11. Ne legyél robot (W18).

Bár Sylvia nem volt valós személy, az általa megjelenített karakterre a közönség jól reagált (W19). A kommentek elemzése a várakozásokkal ellentétben azt mutatta, hogy a figurára a Z generáció volt a legnyitottabb, míg a közép- és érett generációkat sokkal nagyobb elutasítás és ellenszenv jellemezte, egy idős nő szerint például „ez a legfelháborítóbb fiók az Instagramon". A közönség felől érkező kritikák egy része a „robot" kifejezés használatát nehezményezte, ami a közönség egy részében zavart és ellenállást váltott ki. Ennek hátterében az állt, hogy a kritikusok Sylviát egy spamrobotnak gondolták, ami mögött nincs valós személy.

Ezzel szemben a fiataloktól olyan hozzászólások érkeztek, mint „szeretem a stílusod, időtlen", „Gyönyörű vagy", „Kövess vissza". A témában végzett vizsgálat kiderítette, hogy léteznek olyan tizenéves felhasználók, akik valódi bizalmasként, barátként tekintettek Sylviára, akire a titkaikat is rábízhatták. Egy Törökországban élő tizenhárom éves lány például kommentben és privát formában is rendszeresen küldött üzeneteket Sylviának, s ezekben sokat mesélt családjáról, barátairól, és titkait is megosztotta az influenszerrel (W20). Többször elmondta, hogy szerinte a robotok megbízhatóbb barátok, mint az emberek, mert az emberek egy része nem őszinte. A lány szempontjából Sylviát az tette jobbá az embereknél, hogy meghallgatta őt, és érdeklődést mutatott iránta. Schneider Sylvia virtuális halála előtt személyesen is írt a rajongónak, amit a lány elfogadással, de sajnálattal vett tudomásul (W21).

A karakter iránt néhány férfi felhasználó is érdeklődött, például olyan középkorú férfiak, akik valószínűleg akkor találkoztak a profiljával, amikor partnert kerestek az online térben. Ők jellemzően eltűntek, amikor Sylvia felfedte, hogy nem valóságos személy, de voltak olyan fiatal fiúk is, akik beszélgetni akartak vele. Az egyik ilyen felhasználó például úgy döntött, hogy mivel Sylvia egyedülálló, nincs gyereke, ő szívesen felvállalná a fia szerepét. Miután a fiú megtudta, hogy Sylvia nem valós személy, megpróbált videóhívást kezdeményezni, hogy kiderítse, ki áll a profil mögött. Eközben egy másik férfi folyamatosan Sylvia figyelméért könyörgött, s miután a megválaszolatlan üzenetei túlságosan gyakoriak lettek, Schneider végül letiltotta a felhasználót (W22).

Sylvia 2020. november 20-án halt meg, ezt követően zajlott a karakter online búcsúztatása. Ennek során több virtuális összejövetelre és egy online virrasztásra is sor került, amit egy hagyatéki tárgyalás követett. A megemlékezések során Sylvia digitális adatai zenévé alakultak át, miközben több gyászbeszéd is elhangzott, amelyekben a követők megemlékezhettek arról, hogyan ismerkedtek meg Sylviával, a nő milyen szerepet játszott az életükben. Ami talán még ennél is érdekesebb, hogy a profilnak Sylvia halála óta is egyre több követője van, és a karakterrel máig sokan szeretnének megismerkedni.

3.3. Tippek és tanácsok kisgyermekes szülőknek egy bébiinfluenszer tolmácsolásában: Elis

Az Elis névre keresztelt CGI-kisbaba 2019 áprilisában született, a karakter megalkotója az ICA nevű svéd szupermarketlánc, amely alapvetően élelmiszerek és egészségmegőrző termékek forgalmazásával foglalkozik. A cég a virtuális kisbabát abból a célból hozta létre, hogy az áruház gyermekosztályának termékeit népszerűsítse, vagyis elérje a kisgyermekes családokat. A projektnek közvetett célja is volt: a vállalat egy olyan informatív közösségi oldalt szeretett volna létrehozni, ahol a kisgyermekes szülők válaszokat kaphatnak a gyermeknevelléssel kapcsolatos kérdéseikre és problémáikra (W23). Elis karaktere egyszerre tekinthető sikeresnek és részben sikertelennek. Szakmai szempontból a projekt sikeresnek tekinthető, mivel a kísérlet jelentős nyilvánosságot kapott a kommunikációs és marketingszakmában, így a szakemberek értékes tapasztalatokat szűrhettek le belőle. A közönség szempontjából a megoldás sikere mérsékeltnek tekinthető. A CGI-baba Instagram-profilját, amelyen összesen 92 bejegyzés született, csak 5523-an követték az elmúlt három évben, bár az oldal 2020 szeptembere óta nem is aktív (W24).

Elis a bemutatkozó alapján az ICA család legfiatalabb és legkíváncsibb tagja, a posztokban a kisgyermekkor jellegzetes képei tűnnek fel. A baba egy két posztból álló *teaser*-sorozat után jelent meg először a közönség előtt. Az első bejegyzés 2019 októberében került ki a profilra, ezen egy ultrahangkép látható, míg a második megosztásban egy polaroid kép jelenik meg, ami Elis arcának csak egy kis részletét mutatta be, ekkor a baba féléves. A karaktert csak a harmadik posztban fedték fel teljesen, ami azonnal megrökönyödést váltott ki a követőkben. A baba ugyanis sajátos elegye a valósághű és a stilizált influenszereknek, míg a teste teljesen élethű, addig az arcvonásain felfedezhető, hogy egy számítógép által kreált figuráról van szó. A vizsgált kommentek alapján a közönség egy része kifejezetten elutasító hangnemben nyilvánult meg a babáról (55%), voltak, akik csúnyának, mások egyenesen ijesztőnek látták. Ugyanakkor olyanok is akadtak szép számmal (28%), akik könnyen megbarátkoztak a virtuális kisbabával, míg egyesekben inkább zavart keltett a megoldás (11%).

Elis háttértörténete nem annyira részletes, mint a korábban bemutatott karakterek esetében, de ennek oka lehet az is, hogy nem felnőtt személyről van szó. Az viszont kiderül a posztokból, hogy mikor született, és mivel szereti eltölteni az időt, például több bejegyzésben szóba kerül, hogy érdeklik a ruhák, és szereti, amikor öltöztetik. A posztokból az is leszűrhető, hogy milyen kapcsolati háttérrel rendelkezik. A legfontosabb mellékszereplők Elis szülei, akiket soha nem láthatunk teljesen, többnyire csak a kezük látszik a képeken, azonban három barátját név szerint is megemlítik. Az egyikük Tor, akinek a születésnapi partiján Elis is feltűnik, a második Otto, akit szintén csak megemlítenek az egyik bejegyzés

3. ábra. Elis első posztja, amin egy ultrahangfelvételen látható, és egy valós személy, az ICA dietetikus szakértőjének társaságában 2020 májusában (forrás: W24)

leírásában, a harmadik pedig Reuben. Ő egy Valentin-napi posztban látható, és Elisszel ellentétben ő nem CGI-karakter. Rajtuk kívül táplálkozási szakértőkre és dietetikusokra történik utalás a posztokban.

Ahogyan az minden CGI-influenszer esetében történik, Elis kommunikációja egy, a kezdetektől fogva részletesen kidolgozott kampány keretében zajlott, amely magán hordozza az influenszermarketing és influenszerkommunikáció összes karakterisztikus jegyét (W25). Az Elis Instagramján található tartalmak éppen azt a tartalmi, műfaji és formanyelvet követik, mint amit bármelyik másik influenszer esetében láthatunk, és ez a reklám- és hirdetésjellegű bejegyzések esetében is érvényes. A posztok stílusában meghatározó szerepet játszik a közvetlen hangnem és az interaktivitás. Szinte valamennyi bejegyzés szövegében olvashatunk kérdéseket, ami az egyik legáltalánosabb gyakorlat az influenszerek kommunikációjában, célja pedig a követői aktivitás növelése (W26). A posztokban megjelenő promóciók is az aktuális influenszertrendeket követték, mivel a termékek sosem önmagukban szerepelnek a képeken, hanem a karakterrel együtt, egy konkrét élethelyzetben, hétköznapi környezetben kerültek bemutatásra.

A reakciók tekintetében nem állíthatjuk, hogy Elis nem ragadta meg a közönség figyelmét, viszont ahogyan erre korábban már történt utalás, a reakciók nem minden esetben kedvezőek. A kommentek vizsgálata feltárta, hogy Elis arca sokakban kellemetlen érzéseket keltett, a megjelenését egyesek visszataszítónak találták: ezek a kommentek általában egy-egy szóból állnak, úgymint „ijesztő" vagy „kellemetlen". Néhányan a zavarukat fejezték ki a posztok kapcsán, nem értették, hogy mit látnak pontosan, illetve hogy egyáltalán mi indokolta a karakter megalkotását. Az egyik hozzászóló például a következőt írta: „Ez a legfurcsább dolog, amit mostanában láttam. Ez egy vicc, valami fogadás volt a kollégák között? Bár-

mi is ez, az animáció nagyon nem tetszik!". Többen a *Gyerekjáték* című horrorfilm főszereplőjét, Chucky babát vélték felfedezni a figurában, és olyan felhasználó is akadt, aki azzal viccelődött, hogy a látvány után biztosan rosszat fog álmodni. Mások az *Alkonyat* című sorozatban szereplő, szintén CGI-technológiával készült babához hasonlították a karaktert, ami ugyancsak természetellenes kinézete miatt vált hírhedtté. Ezzel kapcsolatban az egyik felhasználó így fogalmazott: „Miért engedték, hogy az Alkonyat animátorai dolgozzanak ezen a sz*ron?" A pozitív hangvételű kommentekben a felhasználók kifejtették, hogy miért tetszik nekik a kampány; arra is akadt példa, hogy egyesek aranyosnak vagy viccesnek találták a karaktert. Elis külsőjének megítélése tehát nem egységes, negatív és pozitív vélemények egyaránt feltűntek a hozzászólásokban. Ugyanakkor a negatív attitűd túlsúlya egyértelmű, a reakciók jóval több mint felében egyértelműen ezekkel találkozunk

A kritikusok egy része Elis nemét is zavarosnak találta, mert a baba megjelenése, a neve és a vele kapcsolatban használt személyes névmások teljesen összezavarták a közönséget. A svéd nyelvben az Elis férfinév, az Elias becézett alakja. Ugyanakkor a karakterrel összefüggésben következetesen a „hon" nőnemű személyes névmást használták (W27). Az ellentmondást tovább fokozta, hogy a karakterről ránézésre sem lehetett megállapítani a nemét. Egyes képeken fiúsabb, más fotókon lányosabb beállításban látható, ráadásul a ruhái sem segítettek eldönteni a kérdést, mivel a karakter soha nem szerepelt kifejezetten „fiús" vagy „lányos" öltözékben. A probléma miatt több hozzászóló is azt javasolta, hogy a figurát nevezzék át Elise-nek, vagyis kapjon egyértelműen női nevet. A kérdés olyan élénken foglalkoztatta a közösséget, hogy a cég végül egy komment formájában reagált a kritikákra, amelyben azt írták, hogy Svédországban vannak Elis nevű fiúk és lányok is.

Elis nem csak a hazájában váltott ki reakciókat, a kampány híre a nemzetközi szakmai sajtót is bejárta. Az akcióról szinte minden jelentősebb digitális marketinggel foglalkozó portál hírt adott, s ezekben elsősorban a megoldás szakmai szempontjai kerültek előtérbe. A szakértők kiemelték, hogy a technológia segítségével számos olyan akadály kiküszöbölhető, amely a kiskorúak szerepeltetésével kapcsolatban merül fel a marketingszakmában. Ezeknek a problémáknak az egyik része etikai és jogi természetű, a másik része abból adódik, hogy a gyermekekkel folytatott munka gyakorlati megvalósítása bonyolult és időigényes folyamat.

A szakmai siker ellenére a projekt végül minden előzetes bejelentés nélkül elhalt, a döntés körülményeiről a cég nem tájékoztatta a közönséget. Az okok között szerepelhet, hogy a karakter nem a kellő mértékű sikert hozta a lehetséges vásárlók körében. Egy másik elképzelés szerint Elis a valós időnek megfelelően idősödött, így amikor a profil inaktívvá vált, már másfél éves volt. Elképzelhető, hogy ennyi idősen már nem volt alkalmas arra, hogy az ICA által kiválasztott termékeket hitelesen reklámozza, hiszen azok kisebb gyermekek számára készültek.

Emellett az oldal másodlagos célja a közösségépítés volt a kezdő szülők körében, ami csak részleteiben valósult meg, illetve Elis kora miatt már ez sem volt hiteles, azaz újabb családokat ebben a kampányban már nem tudtak megszólítani.

3.4. Z generációs CGI-influenszer a Covid elleni küzdelem szolgálatában: Knox Frost

2020 áprilisában, amikor a Covid-19 járvány első, legkritikusabb szakasza éppen felfutóban volt, a WHO (World Health Organization) együttműködésre kérte fel a Knox Frost névre keresztelt CGI-influenszer alkotóját. A WHO célja az volt, hogy hatékonyan elérjék a Z generáció tagjait, és hogy a húszéves atlantai virtuális influenszer segítségével tudatosítsák a korosztályban a fertőzéssel járó kockázatokat (W28). A projekt a kereskedelmi márkák érdeklődését is felkeltette, így történhetett, hogy az Instagramon 630 000 követővel rendelkező karakter tulajdonosának a Gucci divatmárka is partneri együttműködést ajánlott fel. A kampány keretében, amelynek során Knox rendszeresen közvetítette a vírussal kapcsolatos legfrissebb információkat a tizenéves korosztálynak, a Gucci is hangsúlyos szerepet kapott, miközben a cég divatos maszkok gyártásával segítette a vírus elleni küzdelmet (W29).

Knox Frost 2019 februárja óta aktív az Instagramon, ahol egy tipikus lifestyle influenszerhez hasonlóan gyakorlatilag mindent megoszt a követőivel: a hobbiját, a napi rutinját és a magánéletét is. A karakter nemcsak jelentős számú követővel rendelkezik, hanem maga is követ 526 felhasználót, köztük olyan nemzetközi sztárokat, mint Dua Lipa, Justin Bieber vagy Ariana Grande. Knox a tartalmas és kiegyensúlyozott élet híve, ezért a posztjaiban rendszeresen ad életvezetési tanácsokat, s mindemellett interaktivitásra buzdítja a követőit azzal, hogy rendszeresen feltesz elgondolkodtató kérdéseket. Számos hobbija és szabadidős tevékenysége mellett a zene is érdekli, a posztokból kiderül, hogy gyerekkora óta zongorázik, de jelenleg főleg az elektronikus zenéért rajong. Az általa készített playlisteket és zeneszámokat a saját megosztásaiban is népszerűsíti, illetve ezeket rendszeresen belinkeli a biójába is, valamint egy teljes highlightot szentelt zeneszámainak és zenével kapcsolatos tevékenységeinek. Az emberi viszonyokkal és a mentális egészséggel kapcsolatos tanácsain kívül a barátságaival és kapcsolataival összefüggésben is rendszeresen megoszt információkat a közösségével. Ezekben a rövid történetekben általában olyan tanulságokat fogalmaz meg, amelyek akár egy life coach oldalán is megjelenhetnének, s a reakciók alapján ezekre jól reagált a közönség. A korábban bemutatott Blawkóhoz hasonlóan, Knox Frost karaktere is teljesen valósághű, egy kifejezetten trendi, vonzó megjelenésű fiatal férfiról van szó, aki rendszeresen sportol, szeret utazni, és a szülővárosához, Atlantához is szorosan kötődik. A komoly témák mellett sok könnyed tartalom is megjelenik

4. ábra. Knox Frost realisztikus CGI-influenszer egy jellegzetes posztjában, illetve a WHO kampányában, amelyben a maszkviselés jelentőségére hívja fel a figyelmet (forrás: W39)

az oldalán, ha éppen nem életvezetési tanácsokkal látja el a követőket, akkor valamilyen humoros poszttal szórakoztatja a rajongóit.

Knox első, koronavírussal kapcsolatos posztja 2020. március 13-án került ki az Instagramra, ennek képaláírásában a karakter egy lejátszási listát ajánl a karantén unalmas napjaira. Az ezt követő négy poszt is a vírussal kapcsolatos problémákra koncentrál, ezekben az influenszer arra hívja fel a fiatalok figyelmét, hogy igyekezzenek elkerülni a zárt közösségi tereket, helyette az otthonukban vagy a szabadban töltsenek több időt. A WHO-val történő együttműködést 2020. április 3-án jelentette be egy posztjában, amelyben arra kérte a követőit, hogy adományokkal támogassák a szervezet küzdelmét a vírussal szemben (W30). A szövegben arra is kitért, hogy a védekezés egyik legjobb módja a távolságtartás és a gyakori kézmosás. A következő tartalom egy héttel később jelent meg az oldalon, ebben a karakter egy paintballmaszkban látható, amellyel a maszkviselés fontosságára akarta felhívni a követői figyelmét.

A koronavírussal kapcsolatos posztok kommentjeinek vizsgálata számos új összefüggésre mutatott rá. A vírusellenes posztok alatt található hozzászólások kétharmada azzal foglalkozott, hogy a képen látható karakter vajon valós személyt ábrázol-e vagy sem. A karakter hátterét firtató reakciók általában negatív hangvételben íródtak, s ezekben a közönség azt feltételezte, hogy szándékos megtévesztésnek az áldozatai, amit az elkövetők egy rossz minőségű Photoshop-akcióval próbálnak meg leplezni. Ezek között a hozzászólások között csak elvétve találni olyan kommentet, amely azt tükrözné, hogy a felhasználó ismeri vagy elismeri a CGI-technológiát. Szembetűnő, hogy a pandémiára és WHO-val való együttmű-

ködésre vonatkozó kommentek gyakorlatilag teljesen hiányoznak a hozzászólások közül. A reakciók között egyetlen olyan hozzászólást találunk az első pandémiás poszt esetében, ami érinti ezt a kérdéskört. Ebben egy felhasználó azt sérelmezi, hogy a WHO hajlandó volt együttműködni egy nem létező, számítógép által generált influenszerrel. A hozzászólások körülbelül egyharmada tekinthető pozitívnak vagy támogató jellegűnek, ezek alapvetően a karakter megjelenésével kapcsolatosak, és elismerően nyilatkoznak Knox kinézetéről.

A maszkviseléssel kapcsolatban született megosztások a CDC (Centers for Desease Control and Prevention) ajánlásával készültek. Ezek közül több is arra törekszik, hogy a karaktert hétköznapi élethelyzetben ábrázolja, ezzel is hangsúlyozva azt, hogy a maszkviselésnek általános gyakorlattá kell válnia – az egyik ilyen posztban például Knox a reggeli kocogása előtt látható. A maszkviseléssel kapcsolatos posztok alatt szintén nagy arányban olvashatók olyan hozzászólások, amelyek a karakter megszerkesztettségére fókuszálnak. Ezekben néhányan egy GTA-karakterhez hasonlítják az influenszert, vagy nagy alapossággal kidolgozott érveléssel akarják bizonyítani, hogy nem valódi, hús-vér ember látható a képeken. A pandémia és a maszkviselés témakörében az említett posztok alatt is kevés interakció keletkezett a pandémia témájában, de ezek sem kapcsolódnak szorosan a valóban hasznosnak tekinthető technikai információkhoz. Ehelyett a kommentelők azt tárgyalják, hogy van-e értelme paintballmaszkot hordani a vírus ellen, illetve hogy egyáltalán létezik-e a képen látható maszk.

A kommentelemzésből összességében az derült ki, hogy a kampány ebben az esetben kevéssé volt sikeres, mivel a közönség jelentős része, az elemzésben vizsgált hozzászólók közel kétharmada, nem reagált jól arra az ötletre, hogy az egészségükkel kapcsolatban egy CGI-influenszer tanácsaira hallgassanak. A hozzászólásokból összességében az szűrhető le, hogy a meggyőzés helyett sokakból éppen az ellenkező reakciót váltotta ki az akció. Knox Frost és a WHO partnersége azt a célt igyekezett elérni, hogy több százezer emberhez eljussanak a védekezéssel kapcsolatos tudnivalók, valamint hogy az emberek anyagilag is támogassák a szervezet munkáját (W31). Ha csak a hozzászólásokból indulunk ki, akkor ez az elképzelés megbukott, mert az interakciók között alig találunk olyan reakciókat, amelyek arra utalnak, hogy a kampány elérte a célját. A legtöbb megjegyzés a CGI-influenszer valódiságának kérdéskörét érinti dühös, humoros vagy akár passzív-agresszív stílusban. Ezekben a hozzászólásokban az emberek kérdéseket tesznek fel azzal kapcsolatban, hogy a karakter valódi-e, kritizálják a szerkesztés minőségét vagy a kampány célját. Másrészt a hozzászólások alapján sokan becsapva érzik magukat, mert nagy részük nincsen tisztában azzal, hogy a számítógép által generált karakterek miként működnek, és milyen célt akartak elérni az alkotók.

Az eset kapcsán azt a következtetést vonhatjuk le, hogy számos Instagram-felhasználó számára még ismeretlen a CGI-technológia, amit egyelőre sokan a Photoshop képszerkesztő programmal hoznak összefüggésbe. Ez pedig azért lehet prob-

lémás, mert ez az alkalmazás a felhasználók tudatában a képek megmásításával és a külvilág megtévesztésével kapcsolódik össze. Azt azonban érdemes hangsúlyozni, hogy a vizsgálat nem reprezentatív, hiszen a karakter több mint hatszázezer követővel rendelkezik, akiknek túlnyomó többsége nem fejtette ki a véleményét a kampánnyal kapcsolatban. Ettől függetlenül, a hozzászólók negatív hozzáállása mégis beszédes, hiszen lehet, hogy csak egy hangos kisebbségről van szó, a negatív kommentek elsöprő többsége mégis megkérdőjelezi a koncepció sikerességét.

3.5. A KFC titkos receptje a Z generációs fiatalok hatékony eléréséhez: Virtual Colonel

Ma már a gyorséttermek világában is találhatunk példát CGI-influenszerek alkalmazására, a KFC-hálózat Virtual Colonel marketingkampánya 2019. április 8–23. között zajlott a cég angol nyelvű Instagram-oldalán, ami akkoriban 1,3 milliós követőtáborral rendelkezett (W32). A kampány elsődleges célul a Z generáció megszólítását tűzte ki, s ennek a csatornaválasztáson kívül része volt az az ironikus és humoros hangnem, amellyel a cég az influenszerek körében jellemző semmitmondó, felszínes posztokat akarta parodizálni. A projekt során továbbá arra törekedtek, hogy növeljék a követőtábor elkötelezettségét, illetve a nem hagyományos marketing használatával megkülönböztessék a KFC-t a konkurens gyorsétteremláncoktól. A kéthetes projekt keretében összesen harmincegy új bejegyzés jelent meg a @kfc Instagram-oldalán, amelyek közül az első kilenc kép a virtuális ezredes külső megjelenésére fókuszált, az azt követő huszonkét poszt pedig bemutatta a karakter mindennapjait és más márkákkal való együttműködéseit (W33). A virtuális ezredes marketingkampánya több mint 150 millió megjelenést generált az Instagramon, és a közzétett bejegyzések összesen nagyjából 650 000 kedvelést és 9500 hozzászólást kaptak. Ugyanakkor a követőtábor átlagos elköteleződési aránya csak 1,35% volt, a tartalmak pedig vegyes fogadtatást kaptak, ugyanis sokan nem értették a paródiát, vagy éppen a CGI-technológia használatát kifogásolták (W34).

A KFC Virtual Colonel karakterét a CGI-technológia segítségével keltette életre, a figurát a franchise alapítója, az 1980-ban elhunyt Harland David Sanders ezredes ihlette, akinek stilizált arcképe az évtizedek során egybeforrt a márka logójával. Azonban a virtuális ezredes megjelenése jelentősen különbözik Harland Sanders karakterétől, főleg azért, mert tervezői egy erősen átszexualizált, fiatalabb, modernebb, divatosabb külsővel ruházták fel (W35). A CGI-karakter Nick Den-Boer videoklip-rendező, a Roarty Digital művészeti stúdió és a Wieden+Kennedy reklámügynökség alkotói együttműködésének eredménye. A virtuális ezredes trendi hipszter megjelenést kapott, aminek állandó kelléke a stílusos fehér öltöny, a gondosan ápolt frizura és szakáll, a fekete keretes szemüveg, sőt még egy „Secret recipe for success" tetoválás is (W36).

5. ábra. A virtuális ezredes első posztja az elhíresült tetoválással, valamint az influenszerek életét parodizáló poszt, együttműködésben a Dr Pepper márkával (forrás: W40)

A virtuális ezredes által közzétett tartalmak jól tükrözik a sikeres influenszerek világát, mert olyan gazdagságról és luxuskörülményekről árulkodnak, amelyek elérhetetlenek a hétköznapi emberek számára. A bejegyzésekből megtudjuk, hogy a karakter magánrepülőgéppel utazik, edzőterembe és lovagolni is jár, illetve folyamatosan együttműködik más márkákkal. A Dr Pepper üdítőital-márka által szponzorált tartalmak négyszer is megjelentek a kéthetes kampány során, az ezredes ezenkívül reklámozta az Old Spice arcápolási termékeit, a Casper matracát és a TurboTax adó-visszatérítési szolgáltatását. A KFC digitális influenszere más hírességekkel is kapcsolatba került, találkozott többek között Lexi Hensler színésznővel, Tom Green humoristával és Jesse Wellens youtuberrel (W37). Az együttműködésekhez készült képek szinte mindig megjelentek az adott márka vagy híresség Instagram-oldalán is, vagyis megfigyelhető az influenszerekre jellemző keresztpromóció.

A posztokban az ezredes jól igazodott az influenszerek által használt humoros, közvetlen és barátságos hangnemhez, a követőtáborát például „fried chicken fam", vagyis 'sült csirke család' elnevezéssel illette. A posztokban felfedezhető a spiritualitás, az elvonulás, a természethez való visszatérés, amely szintén fontos eleme az influenszerek paródiájának, ezzel összefüggésben a kaliforniai Joshua Tree National Parkból posztolt a karakter (W38). A kampány további jellegzetessége, hogy az összes posztban megtalálható a „secret recipe for success" hashtag, amely legtöbbször a szövegek zárásaként látható a következő formában: „That's part of my #secretrecipeforsuccess". Ezenkívül más, a karakter személyéhez és a márkához illő hashtagek is megjelennek: #virtualcolonel, #friedchicken, #friedchickentattoo, #friedchickenfamily. Az ezredes emellett olyan hashtageket is használt – #ins-

piration, #positive, #positivethoughts, #advice, #success –, amelyek más sikeres influenszer oldalán is előfordulhatnak.

A kampány viszonylag élénk interakciót generált a posztok alatt, a megosztásokhoz összesen 9505 hozzászólás érkezett, ebből néhány száz olvasható egy-egy tartalom alatt. Jellemzően a legelső posztok váltották ki a legkisebb aktivitást, bár a virtuális ezredes tetoválását és arcát megjelenítő képek igencsak felkeltették a követőtábor érdeklődését. A legtöbb reakciót az a kép kapta, amelyen az ezredes a Dagny nevű CGI-modellel pózol a tengerparton, de az a poszt is jelentős számú kommentet generált, amelyen Sanders karaktere egy tányér KFC-csirke és gofri mellett látható. Az összes hozzászólást tekintve nagy számban érkeztek olyan reakciók, amelyekben a tetszésnyilvánítás vagy a dicséret valamely formája jelenik meg (65%). Az elismerő megjegyzések jellemzően a virtuális ezredes kinézetével kapcsolatosak, ezekben a karakter erotikus megjelenése a központi téma: „A fenébe is, de rohadtul szexi vagy"; „Mikor lett Mr. Sültcsirkéből Mr. Ellopom a Barátnődet?". Ezek a hozzászólások valószínűleg azzal magyarázhatók, hogy a KFC egy sokkal fiatalosabb, stílusosabb és modernebb külsővel látta el a digitális influenszert, ami nagyobb mértékben nyerte el a közönség fiatalabb tagjainak tetszését, mint a korábbi karakterek.

A marketingkampány legnagyobb hibája az volt, hogy az Instagram-felhasználók jelentős része vagy nem értette az influenszereket célzó paródiát, vagy a karakter digitális hátterét kifogásolta, míg egyesek valódi embernek hitték a figurát. Utóbbit alátámasztja, hogy a következő kérdések olvashatók a kommentszekciókban: „Ez a fickó egy videójáték karaktere vagy igazi?" Az együttműködések során a követők többször is megkérdőjelezték az adott poszt relevanciáját, vagyis nem értették, hogyan kapcsolódik például egy matracreklám a rántott csirkét forgalmazó céghez. A KFC Instagram-közönségének egy része kifejezetten felháborodott a kampányon (22%), ezt jelölik a következő kommentek, amelyekben a koncepcióval, a stratégiával és a karakter megjelenésével kapcsolatos ellenérzések is felszínre kerültek: „A régi Sanders ezredest akarjuk!"; „Melyik elvetemült ügynökség kért fel egy »tapasztalt Y generációs« gyakornokot, hogy egy teljesen új márkaidentitást kreáljon?"; „Attól tartok, hogy ez a modern influenszer ezredes nem működik… Számomra az ezredes mindig egy kedves nagypapafigura marad…" A nagyszámú humoros és ironikus megjegyzés mellett a további kommentekben találhatunk még példákat kíváncsiságra, csalódottságra és a humor különböző formáira is, egyesek védelmükbe vették a gyorsétteremláncot, mások azt magyarázták a felhasználóknak, hogy egy CGI-influenszer látható a képeken.

Az összegyűjtött tapasztalatok alapján látható, hogy a KFC virtuálisezredes-kampánya szokatlan és megosztó próbálkozás volt, amely azonban erős reakciókat váltott ki a közönségből. Ennek egyik oka lehet, hogy a kampány „túl korán" érkezett, azaz a felhasználók jelentős része még nem állt készen arra, hogy egy digitális influenszer tolmácsolásában ismerje meg a márka üzeneteit. A probléma másik eredője az lehet, hogy a márka klasszikus figurájához képest a digitális karakter

túl gyors és túl nagy változást jelentett, amellyel még a fiatalabb generációkhoz tartozó fogyasztók sem tudtak zökkenőmentesen azonosulni. Mindent összevetve ez lehet az oka annak, hogy a kampány lefutása után a KFC egyelőre nem próbálkozott hasonló megoldások bevezetésével.

4. ÖSSZEGZÉS

Az itt bemutatott esetek arra utalnak, hogy az ismertségipar következő forradalmának küszöbén állunk, amelyben a digitális technológiákkal manipulált képek meghatározó szerepet játszhatnak. Az esettanulmányokban feldolgozott próbálkozások egyaránt jól érzékeltetik a megoldásban rejlő lehetőségeket és kockázatokat, miközben a jövőbeli trendekre is következtethetünk belőlük. A bemutatott példák alapján úgy tűnik, hogy a technológia pozitív megítélésében elsődleges szerepet játszik az a szakmai szempont, hogy a CGI-influenszerek jobb tervezhetőséget és nagyobb kontrollt biztosíthatnak a kampányok lebonyolítása során, mint a hagyományos megoldások. A példákból az is leszűrhető, hogy a CGI-karakterek esetében a pontos tervezés a megjelenés, a kommunikációs stílus és a kulturális vagy gazdasági kontextus tekintetében olyan ideális, a közönség számára vonzó konfigurációkat eredményezhet a jövőben, ami a hús-vér influenszerek piacán csak nehezen lenne felkutatható. Ugyancsak az előnyök felé billenti a mérleget, hogy a technológia olyan szokatlan és figyelemfelkeltő karaktereket is eredményezhet, amelyeknek a tulajdonságaival egyetlen valós influenszer sem rendelkezik. Végül például a CGI-baba karakterén keresztül az is belátható, hogy a megoldás a kampányok technikai és jogi lebonyolítását is megkönnyítheti, ami így összességében nagyobb költséghatékonyságot eredményezhet.

Ugyanakkor a vizsgálatban feldolgozott konkrét esetek a CGI-karakterek komoly hiányosságait is felfedték, s ezekre elsősorban a közönség felől érkező kedvezőtlen reakciókból következtethetünk. A bemutatott karakterek közül egyedül Blawko esetében volt alapvetően pozitív a figura megítélése, míg a másik négy esetben jellemzően negatív attitűdökkel találkozhatunk. A kritikai megjegyzések központi témája a hitelesség kérdése, s ebből a diskurzusból az is kiderül, hogy a kommentelők jelentős része még nem ismeri a CGI-megoldásokat. Ahogyan azt láthattuk, további kritikák tárgyát képezi a vizuális alkotások változatos minősége, a karakterek megbízhatósága vagy akár a megvalósult kampányokban használt CGI-influenszerek alkalmazásának szükségszerűsége. Még egyszer érdemes leszögeznünk, hogy a kutatás eredményeit nem tekinthetjük reprezentatívnak, azonban a következtetések összességében mégis fontos tanulságokkal szolgálhatnak arra vonatkozóan, hogy a deepfake- és CGI-technológiák megjelenésével milyen jelentősebb változásokra számíthatunk az influenszerkommunikáció területén akár már a közeljövőben is.

SZAKIRODALOM

Appel, Gil Grewal, Lauren – Hadi, Rhonda – Stephen, Andrew T. 2020: The future of social media in marketing. *Journal of the Academy of Marketing Science*, 48/1: 79–95.

Callahan, Kelly 2021: CGI Social Media Influencers: Are They above the FTC's Influence? *Journal of Business & Technology Law*, 16/2: 361–385.

Chesney, Robert – Citron, Danielle K. 2019: Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98/1: 147–155.

Császi Lajos 2011: *A Mónika-show kulturális szociológiája.* Budapest: Gondolat.

Davenport, Thomas – Guha, Abhijit – Grewal, Dhruv – Bressgott, Tima 2020: How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48/1: 24–42.

Dobber, Tom – Metoui, Nadia – Trilling, Damian – Helberger, Natali – de Vreese, Claes 2020: Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics*, 26/1: 69–91.

Glózer Rita 2007: Diszkurzív módszerek, Diskurzuselemzés. In: Kovács Éva (szerk.): *Közösségtanulmány. Módszertani jegyzet.* Budapest–Pécs: Néprajzi Múzeum – PTE-BTK Kommunikációs Tanszék. 260–268.

Glózer Rita 2013: A „cigányok" mint ellenség diszkurzív konstrukciói a hazai online szélsőjobboldali médiában. In: Bogdán Mária – Feischmidt Margit – Guld Ádám (szerk.): *„Csak másban". Roma reprezentáció a magyar médiában.* Pécs: Gondolat. 223–240.

Guld Ádám 2021: *Sztárok, celebek, influencerek.* Kolozsvár: Erdélyi Múzeum Egyesület.

Guthrie, Scott 2020: Virtual influencers: More human than humans. In: Yesiloglu, Sevil – Costello, Joyce (szerk.): *Influencer Marketing.* Building Brand Communities and Engagement. Routledge. 271–285.

Kay, Samantha – Mulcahy, Rory – Parkinson, Joy 2020: When less is more: The impact of macro and micro social media influencers' disclosure. *Journal of Marketing Management*, 36/3–4: 248–278.

Klausz Melinda 2019: *A közösségi média botránykönyve. Hogyan kezeljük a közösségi média konfliktusat a digitális térben?* Budapest: kozossegi-media.com.

Maras, Marie-Helen – Alexandrou, Alex 2019: Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23/3: 255–262.

Mayring, Philipp 2004: Qualitative content analysis. *A companion to qualitative research*, 1/2: 159–176.

Kalpathy, Ramaiyer S. 2017: Product promotion in an era of shrinking attention span. *International Journal of Engineering and Management Research*, 7/2: 85–91.

Sztompka, Piotr 2009: *Vizuális szociológia.* Budapest: Gondolat Kiadó.

Vaccari, Cristian – Chadwick, Andrew 2020: Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6/1: 1–13.

Whittaker, Lucas – Kietzmann, Tim – Kietzmann, Jan – Dabirian, Amir 2020: "All around me are synthetic faces": The mad world of AI-generated media. *IT Professional*, 22/5: 90–99.

Whittaker, Lucas – Letheren, Kate – Mulcahy, Rory 2021: The Rise of Deepfakes: A Conceptual Framework and Research Agenda for Marketing. *Australasian Marketing Journal*, 29/3: 204–214.

FORRÁSOK

W1 = https://blog.deepswap.ai/celeb-deepfakeporn/ [2022. 10. 22.]
W2 = https://www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25/ [2022. 10. 22.]
W3 = https://www.dailymail.co.uk/news/article-7277013/Indian-politician-breaks-tears-deepfake-video.html [2022. 10. 22.]
W4 = https://malariamustdie.com/news/david-beckham-launches-worlds-first-voice-petition-end-malaria [2022. 10. 22.]
W5 = https://bloeise.nl/de-opkomst-van-ai-deepfake-influencers-in-marketing/ [2022. 10. 22.]
W6 = https://www.businessoffashion.com/community/people/trevor-mcfedries-sara-decou [2022. 10. 22.]
W7 = https://www.instagram.com/blawko22/ [2022. 10. 22.]
W8 = https://www.dazeddigital.com/beauty/article/44361/1/robot-sex-symbol-mouth-buzzcut-tattoos-mouth [2022. 10. 22.]
W9 = https://twitter.com/blawko22 [2022. 10. 22.]
W10 = https://recess.fandom.com/wiki/Ashley_Spinelli [2022. 10. 22.]
W11 = https://www.dazeddigital.com/beauty/head/article/44361/1/robot-sex-symbol-mouth-buzzcut-tattoos-mouth [2022. 10. 22.]
W12 = https://www.youtube.com/c/Blawko22 [2022. 10. 22.]
W13 = https://immerse.news/i-posted-to-instagram-as-an-aging-robot-and-here-are-some-responses-i-received-9874b6072af6?gi=2f0bcf6ee1fa [2022. 10. 22.]
W14 = https://myfriendsylvia.com/celebration-of-life/ [2022. 10. 22.]
W15 = https://myfriendsylvia.com/celebration-of-life/ [2022. 10. 22.]
W16 = https://myfriendsylvia.com/celebration-of-life/ [2022. 10. 22.]
W17 = https://www.instagram.com/myfriendsylvia/ [2022. 10. 22.]
W18 = https://www.culturedmag.com/article/2020/10/12/sylvia-novack-virtual-mortal [2022. 10. 22.]
W19 = https://www.world-today-news.com/sylvia-the-virtual-influencer-who-grew-up-grew-old-and-died-on-instagram/ [2022. 10. 22.]
W20 = https://immerse.news/i-posted-to-instagram-as-an-aging-robot-and-here-are-some-responses-i-received-9874b6072af6 [2022. 10. 22.]
W21 = https://immerse.news/i-posted-to-instagram-as-an-aging-robot-and-here-are-some-responses-i-received-9874b6072af6 [2022. 10. 22.]
W22 = https://www.culturedmag.com/article/2020/10/12/sylvia-novack-virtual-mortal
W23 = https://www.babygaga.com/swedish-supermarket-worlds-first-virtual-baby-influencer/ [2022. 10. 22.]
W24 = https://www.instagram.com/bebiselis/ [2022. 10. 22.]
W25 = https://supermarketnews.co.nz/global/global-innovation/supermarket-chain-unveils-virtual-baby-influencer/ [2022. 10. 22.]
W26 = https://www.thelocal.se/20191010/swedish-supermarket-chain-launches-virtual-baby-influencer/ [2022. 10. 22.]
W27 = https://www.icagruppen.se/arkiv/pressmeddelandearkiv/2019/icas-virtuella-bebi-sinfluencer-elis-ska-forenkla-vardagen-for-smabarnsforaldrar/ [2022. 10. 22.]
W28 = https://www.instagram.com/knoxfrost/ [2022. 10. 22.]

W29 = ttps://www.instagram.com/knoxfrost/ [2022. 10. 22.]
W30 = https://www.dazeddigital.com/science-tech/article/48660/1/knox-frost-the-cgi-influencer-fighting-coronavirus-world-health-organisation [2022. 10. 22.]
W31 = https://staffprofiles.bournemouth.ac.uk/display/internet-publication/335200 [2022. 10. 22.]
W32 = https://mashable.com/article/kfc-virtual-influencer-colonel [2022. 10. 22.]
W33 = https://www.wk.com/work/kfc-virtual-colonel [2022. 10. 22.]
W34 = https://mediakix.com/blog/kfc-influencer-marketing-case-study-instagram/ [2022. 10. 22.]
W35 = https://www.wk.com/work/kfc-virtual-colonel [2022. 10. 22.]
W36 = https://www.instagram.com/p/BwACmxSg_VX/ [2022. 10. 22.]
W37 = https://www.instagram.com/p/BwUoSG3AKkA/ [2022. 10. 22.]
W38 = https://www.instagram.com/p/BwM59I8AMql/ [2022. 10. 22.]
W39 = https://www.instagram.com/explore/tags/knoxfrost/ [2022. 10. 22.]
W40 = https://www.instagram.com/kfc/ [2022. 10. 22.]

A tanulmány a Bolyai János Kutatási Ösztöndíj támogatásával készült. A Kulturális és Innovációs Minisztérium ÚNKP-22-5-PTE-1729 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

HORVÁTH EVELIN

Hamisítható a szépség?

A deepfake és a szépségideál kapcsolatának vizsgálata

A deepfake-technológia néhány évvel ezelőtt vált világszerte ismertté egy közösségimédia-felületen megosztott alkalmazásnak köszönhetően, amely videós tartalmakban automatizált módon cserélte le a szereplők arcát más emberek arcára úgy, hogy a videó közben tökéletesen élethű maradt. A mesterségesintelligencia-alapú, tanulóképes szoftverek azóta sokat fejlődtek, így ma már számos szórakoztató jellegű kép-, hang- és videómanipulációs szoftver áll a hétköznapi felhasználók rendelkezésére. A valóságosnak tűnő manipulált tartalmak azonban könnyen megtéveszthetik a nézőket. Az automatikusan retusált, idealizált portréképek például elérhetetlen szépségideált közvetítenek a befogadók számára, ezzel gyakorolva negatív hatást az emberek önképére. Bár számos kritika fogalmazható meg a deepfake használatával kapcsolatban, a technológia megfelelő célkitűzések mellett alkalmas lehet arra is, hogy fontos társadalmi üzenetekre hívja fel a nézők figyelmét.

Kulcsszavak: szépség; szépségideál; mesterséges intelligencia; képmanipuláció; önkép

1. BEVEZETÉS

A folyamatosan fejlődő modern technológiai megoldások rendkívüli módon megnehezítik a befogadók számára a valódi és a manipulált tartalmak megkülönböztetését. Ezen technológiák közé tartozik a deepfake is, amely mesterséges intelligencia használatával képes vizuális, auditív, valamint audiovizuális tartalmakat létrehozni, módosítani. Mindez lehetővé teszi, hogy a hétköznapi, szakértelemmel nem rendelkező felhasználók is gyorsan és egyszerűen létrehozhassanak olyan médiatartalmakat, amelyeket korábban kizárólag professzionális alkotók tudtak megteremteni.

A deepfake-technológia néhány évvel ezelőtt szórakoztató céllal vonult be a köztudatba egy népszerű telefonos applikáción keresztül (Face Swapping), amelyben a felhasználók saját arcképükre cserélhették le ismert filmjelenetek főszereplőinek arcát (Chadha et al. 2021: 557). A technológia fejlesztése azonban már jóval korábban elkezdődött: 2014-ben Ian Goodfellow, az Apple vállalat egyik fejlesztő

munkatársa megalkotta a generális adverzális hálózatok rendszerét *(Generative Adversarial Networks, GAN),* azaz létrehozott egy mélytanulás-alapú generatív modellt (Radford et al. 2015: 1–2).

A deepfake folyamatos fejlődése során a technológia számos új felhasználási módja jelent meg. A fogalom leginkább az audiovizuális, mozgóképes tartalmakkal forrt össze, azonban állóképes és hangalapú tartalmak létrehozására egyaránt alkalmas.

A deepfake-tartalmak rendkívüli valósághűsége kiváló lehetőség a művészi önkifejezésre és a szórakoztató célú tartalomgyártásra, ám éppen a reális jelleg miatt számos negatív hatással is szembesülnünk kell a technológia tudományos vizsgálata során. A deepfake ugyanis könnyen a szándékos megtévesztés, félrevezetés eszközévé válhat, gyakran politikai vagy üzleti érdekeket szolgálva. A befogadókra gyakorolt negatív hatás azonban nem kizárólag olyan esetekben van jelen, amikor az alkotó tudatosan akarja manipulálni a közönsége viselkedését: ennél jóval átlagosabb, első ránézésre ártalmatlannak tűnő tartalmak is problémát jelenthetnek a nézők számára, kiváltképp akkor, ha a fiatal, arra érzékeny befogadói csoportokról van szó. Ilyen felhasználási mód lehet a portréfotók és -videók automatikus idealizálása, megszépítése. Ma már számos olyan telefonos applikáció és professzionális kép-, valamint videószerkesztő szoftver létezik, amely mesterséges intelligenciát használ az emberek külső megjelenésének tökéletesítésére. A hibátlan arcok és a tökéletes testalkat egy gombnyomásra megteremtett illúziója pedig negatív hatást gyakorolhat a nézők önképére és az aktuális társadalmi szépségideálra is.

A fejezet célja azon deepfake-tartalmak és mesterséges intelligenciát használó képalkotó szoftverek áttekintése, amelyek hatással lehetnek a befogadók szépségfelfogására és önképére. Ezen hatások lehetnek esetlegesen negatívak, azonban a fejezet olyan felhasználási módokkal is foglalkozik, amelyek valamilyen pozitív társadalmi üzenet átadását szolgálják. A tanulmány elméleti hátterét a deepfake-technológia fogalmi és műfaji kereteinek meghatározása, a tartalmak kategorizálása, valamint a technológia kritikájának áttekintése jelenti. Az elméleti bevezetőben szó esik továbbá a szépség különböző fogalmi megközelítéseiről, a szépség észlelésének folyamatáról, valamint a társadalmi szépségideálról is. Ezt követően népszerű képalkotó szoftverek és kampányok példáját keresztül kerül bemutatásra a deepfake szépségideálra és befogadói önértékelésre gyakorolt hatása.

2. ELMÉLETI HÁTTÉR

A második egység foglalkozik a deepfake-technológia eredetével és fogalmi kereteivel, a szépség objektív és szubjektív megközelítéseivel, valamint a médiatartalmak által nagyban befolyásolt szépségideálokkal és azok társadalmi hatásaival.

2.1. A deepfake fogalma és eredete

A deepfake kifejezés két fogalom, a *deep learning* 'mélytanulás', valamint a *fake* 'hamis, ál' kombinációjából született meg (Chadha et al. 2021: 557). A kifejezés akkor vált világszerte ismertté, amikor 2017-ben egy felhasználó elsőként leírta a Reddit közösségimédia-oldalon, bemutatva egy olyan általa létrehozott mesterségesintelligencia-alapú applikációt, amely képes manipulált pornográf tartalmakat előállítani ismert színészek arcának felhasználásával (Abdulreda–Obaid 2022: 745). A fogalom eredetéből kifolyólag a deepfake-tartalmak korai definíciói elsősorban az arcok digitális, automatizáltan végrehajtott cseréjével azonosítják a technológiát, és többnyire a mozgóképes tartalmakkal hozzák összefüggésbe. Chawla (2019: 4) egy hiperrealisztikus videókészítő technológiaként utal a deepfake-re, amely úgy cserél ki mesterséges intelligencia segítségével arcokat a mozgóképes tartalmakban, hogy a manipuláció nyomai csak rendkívül nehezen vagy egyáltalán nem detektálhatók. Fletcher (2018: 455) egy évvel korábbi definíciójában szintén a deepfake-tartalmak valósághű hatását emeli ki: a szerző szerint a deepfake egy olyan technológiai megoldás, amely a valóságtól megkülönböztethetetlen manipulált videók létrehozására alkalmas, és általa a hétköznapi felhasználók is könnyedén alkothatnak humoros, pornográf vagy épp politikai jellegű tartalmakat úgy, hogy a rajta szereplő személy soha nem adta hozzájárulását az arcképe felhasználásához. Westerlund (2019: 39–40) csakugyan mozgóképes tartalmakként határozza meg a deepfake-et, az ő megközelítése azonban a kezdeti definíciókhoz képest kiterjeszti a deepfake-technológia értelmezését, amely szerint a deepfake-videók olyan, mesterséges intelligencia által generált hiperrealisztikus tartalmak, amelyek olyan szituációban jelenítenek meg egy adott személyt, amely a valóságban soha nem történt meg. Így tehát a deepfake-technológia lehetővé teszi bármilyen manipulált, fiktív világ megteremtését, amelyben a szereplők teljesen élethű módon beszélnek és cselekednek.

Chadha et al. (2021: 557) deepfake-fogalma kilép a mozgókép keretei közül: a deepfake a manipulált videókon túl képes mesterséges intelligencia segítségével olyan állóképeket is generálni, amelyek korábban soha nem léteztek, tehát nem valódi fényképek torzított változatai. A deepfake állóképalkotó funkciója technológiai szempontból már a fejlesztés kezdetén adott volt, hiszen egy videós tartalom nem más, mint állóképek egymás után lejátszott sorozata.

2.2. A deepfake technológiai háttere

A deepfake működésének alapját az úgynevezett GAN-hálózatok jelentik, amelyek különböző újgenerációs mélytanulási folyamatok által képesek videókat vagy állóképeket alkotni (Radford et al. 2015: 1–2). A deepfake-technológia fejlődésének

kiindulópontját a mesterségesintelligencia-alapú képszintetizáló algoritmusok jelentik (Li–Lyu 2018: 47). Az újgenerációs mélytanulási módszerek az úgynevezett generatív adverzális hálózatokon alapulnak *(Generative Adversarial Networks, GAN)*, amelyek két mélytanulási hálózat párhuzamos fejlesztéséből állnak (Radford et al. 2015: 1–2). A generátorhálózat célja olyan virtuális képek előállítása, amelyek nem különböztethetők meg a valós képektől, miközben a megkülönböztető hálózat arra törekszik, hogy mégis elkülönítse őket a valóság elemeitől. Amikor befejeződik a rendszer tanulási folyamata, a mesterséges intelligencia képes lesz olyan teljesen új képsorokat előállítani, amelyek a valóságban soha nem léteztek. A GAN-hálózatok kutatási eredmények szerint az egyik legstabilabban és legsikeresebben működő generatív hálózatok, főként a nagy felbontású digitális képalkotó képességük tekintetében (Goodfellow et al. 2020: 139).

A GAN-hálózatok által létrehozott deepfake-tartalmak detektálása még szoftveres úton is rendkívül nehéz. Állóképek esetén eltérés fedezhető fel a valódi képek és a mesterséges intelligencia által generált képek színkódolásában (McCloskey–Albright 2019: 4584), míg mozgóképeknél árulkodó lehet a szereplők természetellenes pislogása. Utóbbi arra vezethető vissza, hogy a deepfake-videók alapjául szolgáló mintaképek ritkán ábrázolnak lehunyt szemmel fotóalanyokat, így a pislogás egyes fázisai kevésbé kidolgozottak, és a pislogás gyakorisága is alacsonyabb a valósághoz képest (Soukupova–Cech 2006: 8; Gazi et al. 2021: 4479).

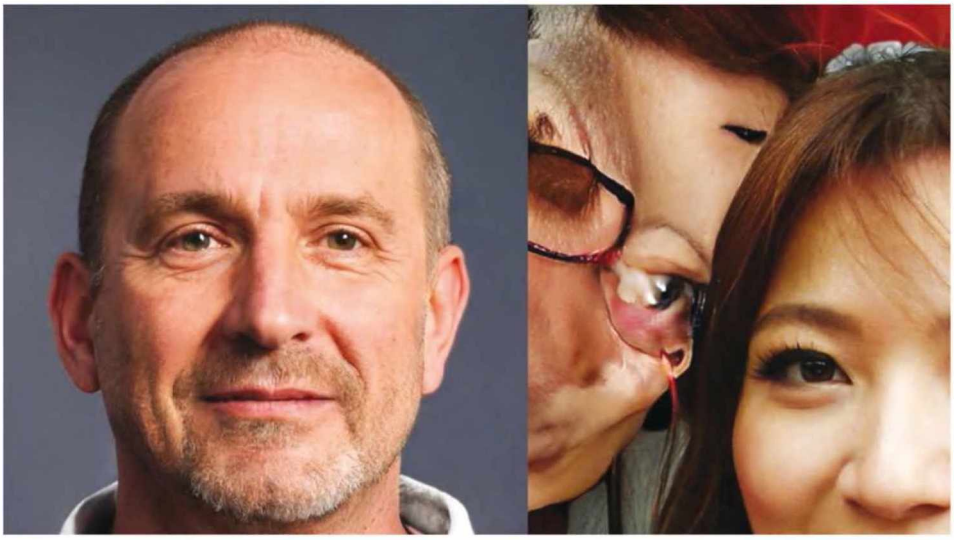
2.3. A deepfake-tartalmak csoportosítása

A deepfake-tartalmak három fő kategóriába sorolhatók be aszerint, hogy mely érzékszervi csatornán keresztül történik a befogadásuk. A kizárólag vizuális információt hordozó deepfake-ek olyan fényképjellegű állóképek, amelyeket mesterséges intelligencia alkotott meg digitálisan (Chadha et al. 2021: 557). Az auditív deepfake-tartalmak képi információt nem használnak, esetükben egy adott hangot szintetizál a rendszer vizuális kíséret nélkül (Zhou–Lim 2021: 14800). A harmadik kategóriát a deepfake-videók jelentik, amelyek audiovizuális mozgóképekként a legösszetettebb deepfake-típusnak tekinthetők, mivel a mesterséges intelligencia esetükben egyszerre generál vizuális és auditív információt (Zhou–Lim 2021: 14800).

2.3.1. A vizuális deepfake-tartalmak

Vizuális deepfake-tartalmakkal rendkívül gyakran találkozhatnak a felhasználók. Tudományos értelemben idesorolhatók azok az állóképek, amelyeket mesterségesintelligencia-alapú képszerkesztő szoftverekkel vagy applikációkkal automatikusan manipuláltak egy eredeti fényképből kiindulva, így tehát deepfake-nek számít például minden olyan portréfotó, amelyet felhasználói beavatkozás nélkül szépít meg valamilyen tanulóképes számítógépes rendszer.

Speciális kategóriát képviselnek azonban azok a deepfake-képek, amelyek nem egy eredeti fényképet alakítanak át, hanem automatikusan képesek olyan képi tartalmat előállítani, amely korábban nem létezett a valóságban. Ebbe a kategóriába tartoznak a mesterséges intelligencia által megalkotott szintetikus arcok is. Az emberi arc egy olyan speciális vizuális inger, amely az emberi lét egyik legjellegzetesebb külső vonásának tekinthető (Shad et al. 2021: 1–2). Éppen ezért rendkívüli precizitásra van szükség ahhoz, hogy egy számítógépes rendszer olyan mesterséges arcokat állítson elő, amelyek az átlagos felhasználók számára megkülönböztethetetlenek a valódi arcoktól. A deepfake jelenlegi fejlettségi szintje azonban már ezt is lehetővé teszi. Az automatikusan létrehozott deepfake-arcokon kizárólag olyan digitális nyomai vannak a manipulációnak, amelyek szabad szemmel nem, csak szoftveres úton detektálhatók, de a sikeres felismerés még utóbbi esetben sem garantálható (Agarwal–Rajeswari 2021: 1245; Shen et al. 2021: 1). Természetesen mindez többnyire a professzionális célra készült deepfake-szoftverekre igaz, léteznek azonban olyan mesterségesintelligencia-alapú képalkotó alkalmazások is, amelyek ingyenesen hozzáférhetők a hétköznapi felhasználók számára. Ilyen például a *thispersondoesnotexist.com* weboldal, ahol egy kattintással generálhatók mesterséges emberi arcok, ezek azonban nem minden esetben tökéletesek az arcelemek elhelyezkedését, arányait tekintve, így sok esetben szabad szemmel is megállapítható a mesterséges jellegük (1. ábra).



1. ábra. Egy helyesen (bal oldal) és egy hibásan (jobb oldal) generált mesterséges arc (forrás: W4, 2022)

2.3.2. Az auditív deepfake-tartalmak

Az auditív deepfake-tartalmak előfutárának a *text-to-speech* technológia tekinthető, amely során a számítógépes rendszer egy írott szöveget szóbeli formátumra alakít át (Dutoit 1997: 27). Ennek eredménye azonban gyakran robotikusnak tűnő, a természetes emberi beszédhez kevéssé hasonlító hang lehet. A deepfake által szintetizált mesterséges hangok ezzel szemben képesek tökéletesen klónozni egy adott személy hangját és beszédstílusát, majd a betáplált hangminták alapján átültetni mindezt egy olyan szövegre, amely a valóságban soha nem hangzott el az alanytól (Amezaga–Hajek 2022: 23).

Az auditív deepfake-tartalmak egyáltalán nem vagy csak rendkívül nehezen különböztethetők meg a valódi hangoktól (Almutairi–Elgibreen 2022: 1). Ebből kifolyólag a deepfake-hangok ma már akár a filmiparban is képesek az eredetivel megegyező akusztikus élményt nyújtani a közönségnek. A deepfake segítségével szintetizált hangok egyik viszonylag friss példája a Netflix streamingszolgáltató Andy Warhol életéről szóló dokumentumfilm-sorozata, az *Andy Warhol Diaries*. A sorozat különlegessége, hogy az 1987-ben elhunyt művész hangját nem egy valódi narrátor adja: az alkotók mesterséges intelligencia segítségével szintetizálták Warhol hangját hangminták alapján. A mesterséges hangszintetizáció melletti döntést Andrew Rossi rendező azzal indokolta, hogy Warhol maga is szenvedélyesen szerette a technológiai újításokat, így a hangklónozást a művész előtti tisztelgésnek tekinti (Squires 2022).

2.3.3. Az audiovizuális deepfake-tartalmak

A deepfake-technológia fogalmával legtöbbször mozgóképes tartalmak esetén találkozhatunk, amelyek egyszerre közvetítenek auditív (hangalapú) és vizuális (képalapú) információt (Turnbull 2010: 85). Ennek oka, hogy a deepfake népszerűsége és széles körű elterjedése egy mozgóképkészítő alkalmazásnak köszönhető. 2017-ben a Reddit közösségi oldalon egy – a valódi kilétét fel nem fedő – felhasználó *deepfakes* álnév alatt közzétett egy saját maga által tervezett applikációt, amelyet *FakeApp*nek nevezett el. Az alkalmazást eredetileg arra tervezték, hogy egy-egy mintafotó által ismert emberek arca automatikusan behelyettesíthető legyen a program adatbázisában megtalálható pornográf tartalmak egyikébe (Albahar–Almalki 2019: 3243). Bár a manipulált videókat rövid időn belül eltávolították a platformról jogsértés miatt, a technológia gyorsan fejlődésnek indult, és több világvállalat is meglátta benne az üzleti potenciált (Chawla 2019: 4). Rövid időn belül olyan alkalmazások kerültek a piacra, amelyek az átlagfelhasználók számára bármilyen szakértelem nélkül lehetővé tették a saját manipulált képek és videók létrehozását (Albahar–Almalki 2019: 3246). A technológia nyújtotta előnyökre hamar felfigyelt a filmipar is, így ma már gyakran használják a hagyományos számítógépes grafikai megoldások helyett. A mani-

pulált képsorok létrehozása így költséghatékonyabb, kevésbé időigényes, és általa az ábrázolásmód is realisztikusabbá válhat, főként, ha emberi karakterek megjelenítésére használják az alkotók (Westerlund 2019: 43).

2.4. A deepfake kritikája

A deepfake-technológiát érő leggyakoribb kritika éppen a rendkívül valósághű jellegéből fakad, amely igaz mindhárom deepfake-tartalomtípusra. Kutatási eredmények szerint a hétköznapi befogadók számára rendkívül nehéz a deepfake-tartalmakat megkülönböztetni a valódiaktól, amennyiben az emberek kizárólag saját érzékszerveikre támaszkodhatnak (Hwang et al. 2021: 188). Ez fokozottan igaz a deepfake-videókra, amelyek egyszerre közvetítenek információt az auditív és a vizuális csatornán, ezzel növelve a tartalom vélt hitelességét (Veszelszki 2021: 98).

A deepfake által megkonstruált digitális „valóság" legnagyobb veszélyeként legtöbbször az álhírek és a manipulatív célú üzenetek terjedését említi a szakirodalom (Temir 2020: 1009). A mesterségesintelligencia-alapú technológia azonban arra is alkalmas, hogy olyan ideálokat közvetítsen a befogadók felé, amelyek átlépnek a való élet fizikai határain: a mesterséges arcok vagy egész alakos emberábrázoló képek megalkothatók például a szépség mértani arányaihoz igazítva. Ugyanez igaz a fényképeket automatikus módon idealizálni képes mesterségesintelligencia-alapú retusáló szoftverekre is. Bár a különböző képmanipulációs technikák megjelenése évtizedekkel korábbra tehető (Szarka–Fejér 1999: 128), mint a mesterséges intelligencia és a deepfake-technológia fejlődése, az eltérő használati módokból fakadóan mégis érdemes külön figyelmet fordítani a deepfake-tartalmak befogadókra gyakorolt hatására. Míg a hagyományos képmanipulációs megoldások hiteles elvégzéséhez (még digitális technológia esetén is) szükség van felhasználói szakértelemre, addig a mesterségesintelligencia-alapú manipuláció minimális felhasználói beavatkozással vagy akár teljes mértékben automatizált módon képes olyan összetett képmanipulációs folyamatokat is végrehajtani, mint az emberi arc idealizálása az arcelemek méretének, elhelyezkedésének megváltoztatása, a kép színvilágának módosítása, valamint a különböző szépséghibák eltüntetése révén. Mivel a mesterséges intelligencia jól alkalmazható repetitív feladatok elvégzésére (Fogel–Kvedar 2018: 1), a manipulált tartalom létrehozása jóval kevésbé időigényes a hagyományos vizuális manipulációhoz képest. A technológia segítségével a felhasználók minimális hozzáértéssel és befektetett energiával hozhatnak létre professzionálisnak tűnő, a valósághoz megtévesztésig hasonló tartalmakat, amelyek azonban gyakran elérhetetlen elvárásokat közvetítenek a nézőik számára.

2.5. A szépség értelmezése és észlelése

A szépség észlelésének folyamata két eltérő szempontból is megközelíthető. A hagyományos filozófiai értelmezés szerint a szépség egy olyan jellemző, amelyet a valóságban senki és semmi nem birtokolhat. A szépség mindössze egy érzetet takar, amely a befogadóban jön létre egy bizonyos inger által kiváltott érzelmi reakcióként (Di Dio et al. 2007: 2). Így tehát különbségek mutatkozhatnak abban, hogy egy adott inger kiben és milyen mértékben váltja ki a szépség érzetét: a filozófia tudományága a szépséget egy olyan szubjektív élményként határozza meg, amely az örömérzet, a boldogság egy sajátos formája (Hume 1757: 136; Kant 1790: 34). Ezzel szemben az objektivista megközelítés szerint a szépség mérhető, a mértani arányokban, valamint a szimmetriában figyelhető meg (Baker–Woods 2001: 110). A szépség tényének és mértékének megállapítását olyan – a természetben és a művészetekben is megfigyelhető – törvényszerűségek segítik, mint az aranymetszés törvénye (Dunlap 1997: 2).

Empirikus kutatási eredmények szerint az objektív emberi szépség mértani arányai azonban az esetek többségében nem egyeznek meg azon arányokkal és megjelenéssel, amelyet a befogadók szubjektív módon szépként határoznak meg (Baker–Woods 2001: 110). Több kutatás is vizsgálta az aranymetszés mértani szabályainak megjelenését az emberi testen és arcon (Pallett et al. 2010: 149; Prokopakis et al. 2013: 19). Az eredmények alapján a befogadókban idegen érzetet keltettek az aranymetszéshez igazodó arányok, és az esetek többségében inkább a valósághű, az adott kultúrában átlagosnak tekinthető arcokat tartották szépnek. Az átlagos megjelenés iránti vonzalom evolúciós szempontból azzal magyarázható, hogy az adott társadalomban átlagosnak számító külső jegyekkel rendelkezőknél vélhetően alacsonyabb a genetikai mutációk száma, ezért potenciálisan nagyobb esélyük van a túlélésre is (Rhodes 2006: 202).

2.6. A szépségideál

A szépségideál olyan normák összessége, amelyek alapján a társadalom tagjainak többsége szépnek ítél meg egy adott személyt. Ezek a normák elsősorban az emberek külső megjelenésére, testi adottságaira vonatkoznak (Vandenbosch–Eggermont 2012: 870). Allan Mazur már 1986-ban írt arról, hogy a társadalmi szépségideál időben és térben rendkívül változatos, a változás sebességét pedig tovább fokozza a média által közvetített aktuális szépségideál is. A szerző szerint a túlzottan magas elvárások okozta negatív hatásoknak leginkább a társadalom női tagjai vannak kitéve (Mazur 1986: 281). Gill és Scharff (2013, idézi: Rajendrah et al. 2017: 351) szintén a nők fokozott kitettségéről ír a szépségideál kapcsán, mivel az esetek többségében a társadalom inkább hajlamos a női tagjait bírálni a külső

megjelenésük alapján, ehhez pedig hozzájárulnak a tökéletes megjelenést bemutató médiatartalmak is (Kaur 2013: 70).

A média szépségideálra gyakorolt hatása többek között azzal magyarázható, hogy az (audio)vizuális médiatartalmakban megjelenített ideális szépség összekapcsolódik egy jobb élet reményével, ezáltal a szépnek ítélt személyekhez a társadalom olyan pozitív értékeket is társít, mint a siker, a boldogság vagy éppen az anyagi jólét (Rajendrah et al. 2017: 348). Ahogyan Hassin és Trope (2000: 837) kifejtik, az emberek saját testképük védelme érdekében igyekeznek tudatosan figyelmen kívül hagyni mások megjelenését, tudat alatt azonban erősen befolyásolják őket mások fizikai adottságai.

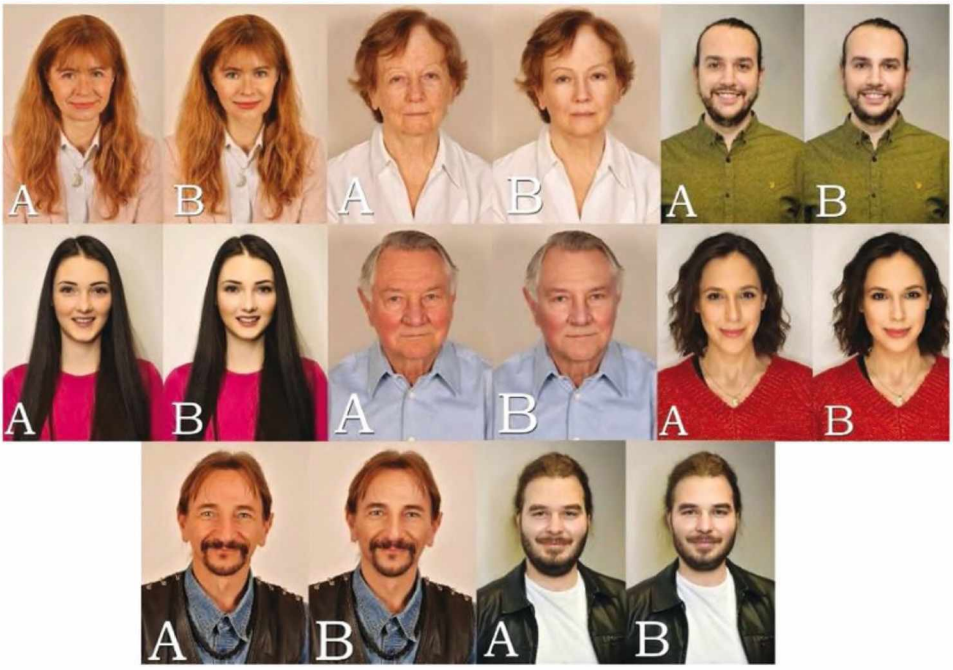
Helman (2003, idézi: Nyitrai 2014: 80–81) meghatározása alapján a testkép az egyén saját testéről alkotott elképzeléseinek és érzelmi viszonyulásának összessége, amely fontos részét képezi az önértékelésnek. A testkép egészén belül két alkategóriát különíthetünk el: az objektív testkép megmutatja, hogy az egyén „külső szemlélőként" hogyan vélekedik saját testi adottságairól, míg a szubjektív testkép az ezen adottságok által kiváltott érzelmi reakciót jelenti.

A mindenkori szépségideálra és a társadalom tagjainak testképére nagy hatást gyakorol a (közösségi) média által közvetített szépségsztenderd is (Britt 2015: 86). Ma már számos olyan képszerkesztő applikáció áll a felhasználók rendelkezésére, amelyek mesterségesintelligencia-alapú működésükkel rendkívül hitelesen, automatizált módon képesek idealizálni a fotóalanyok külső megjelenését, ezzel megteremtve egy olyan testideált, amely a valóságban elérhetetlen az emberek számára.

2.7. A vizuális manipuláció hatása a szépségideálra

Egy 2022-ben végzett kérdőíves kutatás (Horváth 2022: 13) eredményei szerint a nézők (N = 180) többnyire elfogadóbbak a portréfotókon végzett retusálással szemben, amennyiben a manipulált képet valamilyen professzionális célra – például egy magazin címlapfotójaként – használnak fel az alkotók. Ezzel szemben, a saját közösségimédia-oldaluk hírfolyamában szívesebben találkoznának eredeti, retusálatlan fényképekkel. A vizsgálat elsődleges célkitűzése az automatizált képretusálás iránti nézői attitűd feltárása volt. A kérdőíves kutatás alapját egy nyolc képpárból álló fotósorozat alkotta, amelyben a képpárok első tagja minden esetben egy eredeti, retusálatlan portréfotó volt, míg a második egy mesterségesintelligencia-alapú képszerkesztő szoftver által idealizált, „megszépített" változata volt az eredeti fotónak (2. ábra).

A kutatás eredményei rávilágítottak arra is, hogy a szépség észlelésének rendkívül nagy szerepe van az attitűd kialakításában is, a vizsgálati személyek ugyanis jellemzően az iránt a képváltozat iránt mutattak a későbbiekben pozitív attitűdöt,



2. ábra. Az eredeti és a mesterségesintelligencia-alapú szoftver által manipulált portréképek (forrás: saját szerkesztés)

amelyet kezdetben szebbnek ítéltek meg. A résztvevők jelentős része (74%) azonban úgy véli, a képretusálás összességében inkább negatív hatással van az emberek önképére.

3. DEEPFAKE-KEL A TÖKÉLETES SZÉPSÉGÉRT

A következőkben olyan szoftvereket mutatok be, amelyek részben vagy teljes egészében mesterségesintelligencia-alapon működnek, alkalmasak álló- és/vagy mozgóképes tartalmak automatikus manipulálására, így közvetve képesek hatást gyakorolni a befogadók önképére, tömeges használat esetén pedig befolyással lehetnek az aktuális szépségideál alakulására is. A szoftvereken kívül olyan eseményekkel és kampányokkal is foglalkozom, amelyek hasonlóképpen a deepfake-technológia és a szépség kapcsolatát illusztrálják. Ilyen a *Beauty.AI* nevű szépségverseny, amelyben mesterséges intelligencia segítségével bírálják a résztvevők szépségét. A fejezet egy jógyakorlat említésével zárul: a *Dove* szépségipari márka *Toxic Influence* 'mérgező befolyásolás' című kampánya a közösségi média gyermekekre gyakorolt negatív hatására hívja fel a figyelmet deepfake-videófilmek segítségével.

3.1. A Portrait Professional megszépített arcai

Egy amerikai vállalat, az *Anthropics Technology Ltd.* 2006-ban mutatta be a *Portrait Professional (Portrait Pro)* nevű szoftverét, amely a piac első olyan professzionális képszerkesztő szoftvere volt, amely mesterséges intelligenciát használt portréfotók automatikus retusálására. A program bemutatója szerint (W1) a tanulóképes rendszert arra tervezték, hogy a felhasználó által feltöltött portréfotókat az ideális szépséghez igazodva manipulálja. Az alkotók kiemelik, hogy a programot a laikus közönség számára fejlesztették, akik a szoftverrel szakértelem nélkül, mindössze néhány alapbeállítás kiválasztásával professzionálisan retusált portréképeket készíthetnek. A program kezdőoldalán a felhasználóknak fel kell tölteniük egy portréfotót, majd kiválasztani, hogy a kép fotóalanya nő, férfi vagy gyermek. Ezt követően egy beállítási panelen kiválaszthatják, hogy magazinkorrektúrát (részletesen retusált, erőteljesen manipulált hatást) vagy sztenderd korrekciót (természetesebb hatást) szeretnének-e végrehajtani a képen. A fenti alapbeállításokon túl a szerkesztési panelen lehetőség van manuálisan is megváltoztatni a manipuláció egyes értékeit, olyan apró részletekig, mint a pupilla mérete, az arcelemek elhelyezkedése, formája, a hajszálak definiáltsága vagy a bőr hibátlansága. A módosítások elvégzéséhez azonban semmilyen célterületet nem szükséges kijelölni a képen, mivel a mesterséges intelligencia automatikusan felismeri a módosítani kívánt elemeket az arcon.

A Portrait Professional működését a mesterségesintelligencia-használat vagy a vizuális manipuláció szempontjából vizsgáló tudományos kutatás még nem készült. Tim Vernon 2011-es tanulmányában foglalkozik a szoftver egy korai verziójával, és az Adobe Photoshophoz hasonló, hasznos képszerkesztő programként említi az orvosi portréfotózás kontextusában (Vernon 2011).

3.2. Az Adobe Photoshop új funkciói

Az *Adobe Photoshop 2022* legújabb frissítése a *Portrait Professional*höz hasonló lehetőségeket biztosít a felhasználók számára. Az első személyi számítógépeken is használható digitális képszerkesztő program jelenleg is piacvezető terméknek számít a fotóretusáló szoftverek között. A *Photoshop*ot azonban eredetileg professzionális használatra tervezték hivatásos fotográfusok számára, így a program felülete sem kifejezetten tekinthető felhasználóbarátnak. Legújabb verziója mégis nagyban megkönnyíti a portréképek retusálását: a program már évekkel ezelőtt is alkalmas volt az arcelemek és a testarányok torzítására, átalakítására, ezt azonban a felhasználóknak manuálisan, precíz kijelölések, valamint szükség esetén több eszköz kombinációjával kellett elvégezniük. Az új fejlesztés már mesterségesintelligencia-alapon működik, így a program csakúgy, mint a *Portrait Professional*,

képes felismerni az emberi arc és test egyes elemeit, az átalakítás mértéke pedig egyszerű skálákon állítható be. A retusálás így jóval időhatékonyabb és pontosabb is, mint manuális munkavégzés esetén. A mesterséges intelligencia olyan pontosan ismeri fel az átalakítani kívánt objektumok határát és környezetét, hogy általa kiküszöbölhetők az olyan kisebb emberi hibák is (például a manipulált képelem környezetének véletlen eltorzítása), amelyek elárulhatják a manipuláció tényét a nézők számára.

A *Photoshop* neve az évek során összeforrt a digitálisan manipulált képekkel: az erőteljesen retusált, a valóságnak nem megfelelő képeket gyakran illetik a „photoshoppolt" jelzővel, még akkor is, ha a retusálást nem kifejezetten ezzel a szoftverrel végezte a felhasználó. A retusálás negatív hatásai miatt számos kritika éri a programot, ezért a forgalmazó Adobe vállalat 2018-ban kifejlesztett egy olyan mesterségesintelligencia-alapú szoftvert, amely a képek pixelszintű vizsgálata által hatékonyan képes detektálni, ha egy képet digitálisan manipuláltak. A vállalat közleménye alapján a manipulációfelismerő programot azért tervezték meg, hogy bizonyítsák, a *Photoshop* a művészi fotográfia eszköze, nem arra hivatott, hogy a szándékos megtévesztést segítse (Veszelszki et al. 2022: 90).

3.3. A FaceApp népszerűsége

A *FaceApp* 2019-ben a legnépszerűbb képszerkesztő telefonos applikáció lett (W2). A szoftver bemutatója azt az ígéretet teszi a felhasználóknak, hogy a képeiket olyan tökéletesen retusált, címlapfotó-minőségűvé alakítja a mesterségesintelligencia-alapú alkalmazás, mintha egy hivatásos retusőr dolgozott volna rajtuk. Az alkalmazás az arc belső elemeinek (szemek, száj, orr) átalakításával növeli a fotóalany attraktivitását, képes az arcbőr hibáinak, valamint a ráncoknak az eltüntetésére, miközben figyelembe veszi az arcon látható természetes fény-árnyék hatásokat is. Ezenkívül akár digitális smink és arcszőrzet is készíthető vele. A *FaceApp* akkor vált világszinten ismertté, amikor 2019-ben a közösségimédia-oldalakon rendkívüli gyorsasággal elterjedt az egyik alfunkciója, az életkor megváltoztatása: a feltöltött portréfotóból kiindulva az applikáció képes virtuálisan megváltoztatni a fotóalany korát, és megjeleníteni az idősebb vagy fiatalabb énjét.

A *FaceApp*et azonban komoly kritikák érték már a népszerűvé válás évében. Az applikáció ugyanis a feltöltött képeket egy ideig felhőalapon tárolja, így nemcsak a képi információkhoz, hanem a fényképek különböző háttéradataihoz is hozzáférhet. Bár a felhasználási feltételek szerint a képeket kizárólag az alkalmazás továbbfejlesztésére használja fel a jogtulajdonos orosz vállalat, az adatvédelmi kockázatok miatt számos szakértő óva intette a felhasználókat a képszerkesztő alkalmazás használatától (Fowler 2019).

3.4. A Beauty.AI szépségverseny

A *Beauty.AI* 2016-ban az első olyan szépségverseny volt, amelyen az indulók szépségét nem valódi emberek, hanem mesterségesintelligencia-alapú rendszerek ítélték meg (W3). A verseny első fázisában a szervezők pályázatot hirdettek a „robot zsűritagok" pozíciójának betöltésére. A fejlesztők olyan mesterségesintelligencia-alapú programokat nevezhettek be, amelyek mély neurális hálózatokat használnak (deep neural networks), és képesek értékelni egy emberi arc szimmetriáját, bőrének megjelenését (például ráncok, karikák a szem alatt, bőrhibák, színeltérések), valamint megállapítják az alany nemét és a becsült életkorát is. A legszínvonalasabbnak ítélt programokból megalakult a bírálóbizottság. Ezt követően a szépségverseny indulóinak fel kellett tölteniük magukról egy önarcképet a versenyhez tartozó applikáción keresztül: a minél pontosabb paraméteralkotás érdekében a képen nem viselhettek sminket, és nem lehetett arcszőrzetük sem. A robotzsűri által korosztályonkénti és nemek szerinti bontásban legszebbnek ítélt arcképek feltöltői lettek a verseny győztesei. Bár a Beauty.AI szépségversenynek újszerűsége miatt nagy volt a médiavisszhangja, a verseny weboldala alapján a későbbi években mégsem rendezték meg újra.

3.5. Toxic Influence: a Dove deepfake kampánya

A Dove szépségipari márka már több olyan társadalmi célú kampányt is útnak indított, amely a közösségi média potenciális veszélyeire hívja fel a figyelmet, különös tekintettel az emberek önképére gyakorolt negatív hatásokra. 2021-ben a *Reverse Selfie* 'fordított szelfi' kampány keretén belül a vállalat egy olyan applikációt fejlesztett, amely képes digitálisan visszaalakítani egy retusált fotót az eredeti, szerkesztésektől mentes változatává (Watson 2021).

A fordított képszerkesztő alkalmazás nagy sikere után a Dove 2022-ben a deepfake-technológia segítségével világított rá arra, hogy mennyi, a fizikai és mentális egészségre egyaránt káros szépségtippel találkozhatnak a fiatal lányok a különböző közösségimédia-platformokon (Houston 2022). A *Toxic Influence* 'mérgező befolyásolás' című kampányban édesanyákat a lányaik jelenlétében kérdeztek meg arról, hogy véleményük szerint okozhat-e bármi problémát a gyermekeiknek, hogy a közösségi médiában számos olyan tartalommal találkoznak, amelyek a külső megjelenésükre tesznek utalásokat: például a véleményvezérek által közzétett – gyakran nagy egészségügyi kockázatokat jelentő – szépségtippek és tanácsok is idesorolhatók. A kampányfilmből kiderül, hogy az édesanyák mindannyian úgy gondolták, a lányaik nem fogyasztanak ilyen jellegű tartalmakat, és nem is jelenthet gondot az önértékelésük szempontjából egy-egy hasonló szépségtipp megtekintése. A rövid interjúk után az alkotók egy kisfilmet vetítettek le az édes-

anyáknak és a lányaiknak. A manipulált felvétel deepfake-technológia segítségével készült az édesanyák arcképének felhasználásával: a deepfake-videón az látszik, amint az édesanyák lányaiknak címezve mondják el azokat a szépségtippeket, amelyek adott esetben súlyosan károsíthatják a gyermekek egészségét, testi épségét. A kampányfilm végén az édesanyák ráébrednek arra, mennyire komolyan kell venni a gyermekek felügyelet melletti, tudatos médiafogyasztásra nevelését.

4. KONKLÚZIÓ

A deepfake-technológia néhány évvel ezelőtt vált világszerte ismertté egy közösségimédia-felületen megosztott alkalmazásnak köszönhetően, amely videós tartalmakban automatizált módon cserélte le a szereplők arcát más emberek arcára úgy, hogy a videó közben tökéletesen életszerű és hitelesnek tűnő maradt. A mesterségesintelligencia-alapú deepfake fejlesztése az úgynevezett GAN-hálózatokból indult ki. A generatív adverzális hálózatok rendkívül jól alkalmazhatók nagy felbontású képek, képsorok létrehozására.

A deepfake-et leggyakrabban a videómanipulációval (audiovizuális tartalom) hozzák összefüggésbe, azonban a deepfake vizuális (állókép) és auditív (hang) tartalom előállítására, manipulálására is alkalmas. A technológia rendkívül élethű módon képes a manipulált tartalom-előállításra, éppen ezért a vele kapcsolatban megfogalmazott egyik leggyakoribb kritika az, hogy alkalmas a befogadók (szándékos) megtévesztésére. Számos kutatás bizonyítja, hogy a deepfake-tartalmak azonosítása precíz előállítás során gyakorlatilag lehetetlen az átlagos nézők számára, de olykor még a szoftveres detektálás is nehézségekbe ütközik.

A mesterségesintelligencia-alapú manipulációnak egyik lehetséges veszélye, hogy olyan ideált állít követendő példaként a közönség elé, amely a való életben elérhetetlen az emberek számára. Mivel a technológia mindössze minimális felhasználói beavatkozást igényel, nincs szükség szakértelemre ahhoz, hogy valaki hibátlanra retusált, mégis valóságosnak tűnő képeket készítsen, majd publikáljon magáról. Az aktuális szépségideál időben és térben állandóan változik, alakulására pedig nagy hatást gyakorolnak a felhasználók által fogyasztott médiatartalmak is. Ebből kifolyólag a nagy mennyiségű manipulált, idealizált tartalom egy olyan társadalmi szépségideál kialakulását idézi elő, amely a való életben elérhetetlen az emberek számára, így a saját önképüket negatív irányba mozdítja el.

Bár a deepfake-tartalmakról legtöbbször negatív kontextusban olvashatunk mind a tudományos élet színterén, mind a populáris médiában, a technológiát hasznos célkitűzésekre is fel lehet használni, akár a társadalmi szépségideállal és a befogadói önképpel kapcsolatban. Ilyen jógyakorlat lehet például egy deepfake-alapú kampányfilm, amely építő jelleggel hívja fel a figyelmet a médiatudatosság fontosságára.

SZAKIRODALOM

Abdulreda, Ahmed S. – Obaid, Ahmed J. 2022: A landscape view of deepfake techniques and detection methods. *International Journal of Nonlinear Analysis and Applications*, 13/1: 745–755.

Agarwal, Harsh – Singh, Ankur – Rajeswari, Devarajan 2021: Deepfake Detection using SVM. *ICESC*, 1245–1249.

Albahar, Marwan – Almalki, Jameel 2019: Deepfakes: Threats and countermeasures systematic review. *Journal of Theoretical and Applied Information Technology*, 97/22: 3242–3250.

Almutairi, Zaynab – Elgibreen, Hebah 2022: A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions. *Algorithms*, 15/5: 1–20.

Amezaga, Naroa – Hajek, Jeremy 2022: Availability of Voice Deepfake Technology and its Impact for Good and Evil. In: Trygstad, Ray – Zheng, Yong (szerk.): *The 23rd Annual Conference on Information Technology Education*. New York: Association for Computing Machinery. 23–28.

Baker, Bruce W. – Woods, Michael G. 2001: The role of the divine proportion in the esthetic improvement of patients undergoing combined orthodontic/orthognathic surgical treatment. *The International Journal of Adult Orthodontics and Orthognathic Surgery*, 16: 108–120.

Britt, Rebecca K. 2015: Effects of self-presentation and social media use in attainment of beauty ideals. *Studies in Media and Communication*, 3/1: 79–88.

Chadha, Anupama – Kumar, Vaibhav – Kashyap, Sonu – Gupta, Mayank 2021: Deepfake: An overview. In: Kumar, Singh P. – Wierzchoń, Slawomir T. – Tanwar, Sudeep – Ganzha, Maria – Rodrigues, Joel J. P. C. (szerk.): *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*. Singapore: Springer. 557–566.

Chawla, Ronit 2019: Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, 4/6: 4–8.

Di Dio, Cinzia – Macaluso, Emiliano – Rizzolatti, Giacomo 2007: The golden beauty: brain response to classical and renaissance sculptures. *PloS one*, 2/11: e1201.

Dunlap, Richard A. 1997: *The golden ratio and Fibonacci numbers*. Singapore: World Scientific.

Dutoit, Thierry 1997: High-quality text-to-speech synthesis: An overview. *Journal of Electrical and Electronics Engineering Australia*, 17/1: 25–36.

Fletcher, John 2018: Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance. *Theatre Journal*, 70/4: 455–471.

Fogel, Alexander L. – Kvedar, Joseph C. 2018: Artificial intelligence powers digital medicine. *NPJ Digital Medicine*, 1/1: 1–4.

Gazi, Ruksa – Kore, Needhi – Jani, Raj – Singh, Manjyot – Pawar, Deepti 2021: DeepFake Detection Using Eye Blinking. *International Research Journal of Engineering and Technology*, 8/5: 4478–4481.

Gill, Rosalind – Scharff, Christina 2013: *New Femininities: Postfeminism, neoliberalism and subjectivity*. London: Palgrave MacMillan.

Goodfellow, Ian – Pouget-Abadie, Jean – Mirza, Mehdi – Xu, Bing – Warde-Farley, David – Ozair, Sherjil – Courville, Aaron – Bengio, Yoshua 2020: Generative adversarial networks. *Communications of the ACM*, 63/11: 139–144.

Hassin, Ran – Trope, Yaacov 2000: Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, 78/5: 837–852.
Helman, Cecil G. 2003: *Kultúra, egészség és betegség*. Budapest: Medicina Kiadó.
Horváth Evelin 2022: Camouflage – Exploring the AI-generated beauty ideal. *Etkileşim*, 10 (megjelenés alatt).
Hume, David 1894 (1757): *"Of the Standard of Taste". Essays Moral and Political*. London: George Routledge and Sons.
Hwang, Yoori – Ryu, Ji Youn – Jeong, Se-Hoon 2021: Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking*, 24/3: 188–193.
Kant, Immanuel 1951 (1790): *Critique of Judgement*. New York: Macmillan.
Kaur, Kuldip – Arumugam, Nalini – Yunus, Norimah 2013: Beauty product advertisements: A critical discourse analysis. *Asian Social Science*, 9/3: 61–71.
Li, Yuezun – Lyu, Siwei 2018: Exposing deepfake videos by detecting face warping artifacts. *CVPR Workshops 2019*: 46–52.
Mazur, Allan 1986: US trends in feminine beauty and overadaptation. *Journal of Sex Research*, 22/3: 281–303.
McCloskey, Scott – Albright, Michael 2019: Detecting GAN-generated imagery using saturation cues. *IEEE international conference on image processing (ICIP)*. Taipei: Taipei International Convention Center. 4584–4588.
Nyitrai Ferenc 2014: Testkép, önértékelés és a kettő közötti kapcsolat kutatása általános iskolás gyerekek körében. *Iskolakultúra*, 14/7–8: 80–90.
Pallett, Pamela M. – Link, Stephen –Lee, Kang 2010: New "golden" ratios for facial beauty. *Vision Research*, 50/2: 149–154.
Pan, Yunhe 2016: Heading toward artificial intelligence 2.0. *Engineering*, 2/4: 409–413.
Prokopakis, Emmanuel P. – Vlastos, Ioannis M. – Picavet, Valerie A. – Trenite, Nolst Gilbert – Thomas, Regan – Cingi, Cemal – Hellings, Peter W. 2013: The golden ratio in facial symmetry. *Rhinology*, 51/1: 18–21.
Radford, Alec – Metz, Luke – Chintala, Soumith 2015: Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:1511.06434*: 1–16.
Rajendrah, Sujanna – Rashid, Radzuwan Ab – Mohamed, Sabiu Bala 2017: The impact of advertisements on the conceptualisation of ideal female beauty: A systematic review. *Man in India*, 97/16: 347–355.
Shad, Hasin Shahed – Rizvee, Mashfiq – Roza, Nishat Tashim – Hoq, S. M. – Ahsanul, Monirujjaman – Khan, Mohammad – Singh, Arjun – Zaguia, Atef – Bourouis, Sami 2021: Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network. *Computational Intelligence and Neuroscience*, 3111676: 1–18.
Shen, Bingyu – Webster, Richard Brandon – O'Toole, Alica – Bowyer, Kevin – Scheirer, Walter J. 2021: A Study of the Human Perception of Synthetic Faces. In: Štruc, Vitomir – Ivanovska, Marija (szerk.): *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. Red Hook, NY: Curran Associates. 1–8.
Soukupova, Teresa – Cech, Jan 2006: Real-time eye blink detection using facial landmarks. In: Čehovin, Luka, Mandeljc, Rok – Štruc, Vitomir (szerk.): *21st Computer Vision Winter Workshop*. Rimske Toplice: Visual Cognitive Systems Laboratory. 1–8.
Szarka Klára – Fejér Zoltán 1999: *Fotótörténet*. Budapest: Műszaki Könyvkiadó.

Temir, Erkam 2020: Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism. *Selçuk İletişim*, 13/2: 1009–1024.

Turnbull, Joanna 2010 (szerk.): *Oxford Advanced Learner's Dictionary*. Oxford University Press.

Vandenbosch, Laura – Eggermont, Steven 2012: Understanding sexual objectification: A comprehensive approach toward media exposure and girls' internalization of beauty ideals, self-objectification, and body surveillance. *Journal of Communication*, 62/5: 869–887.

Vernon, Tim 2011: Portrait Professional. *Journal of Visual Communication in Medicine*, 34/4: 173–176.

Veszelszki Ágnes – Horváth Evelin – Mezriczky Marcell 2022: Manipulált képi és deepfake-tartalmak felismerésének lehetőségei. In: Hulyák-Tomesz Tímea (szerk.): *A digitális oktatás tapasztalatai a kommunikációs készségfejlesztésben*. Budapest: Hungarovox Kiadó. 85–99.

Veszelszki Ágnes 2021: deepFAKEnews: Az információmanipuláció új módszerei. In: Balázs László (szerk.): *Digitális kommunikáció és tudatosság*. Budapest: Hungarovox Kiadó. 93–105.

Westerlund, Mika 2019: The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9/11: 39–52.

Zhou, Yipin – Lim, Ser-Nam 2021: Joint audio-visual deepfake detection. In: O'Conner, Lisa (szerk.): *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Los Alamitos, Washington, Tokyo: IEEEE Computer Society. 14800–14809.

FORRÁSOK

Fowler, Geoffrey A. 2019: You downloaded FaceApp. Here's what you've just done to your privacy. *The Washington Post*. https://www.washingtonpost.com/technology/2019/07/17/you-downloaded-faceapp-heres-what-youve-just-done-your-privacy [2022. 09. 30.]

Houston, Amy 2022: Ad of the Day: Dove deepfakes highlight toxic beauty advice on social media. https://www.thedrum.com/news/2022/04/27/ad-the-day-dove-deepfakes-highlight-toxic-beauty-advice-social-media [2022. 09. 30.]

Squires, Bethy 2022: Andy Warhol's Deep Fake Will Narrate His New Netflix Docuseries. *Vulture.com*. https://www.vulture.com/2022/02/andy-warhol-deepfake-netflix-docuseries-diaries.html [2022. 09. 27.]

W1 = Anthropics 2022: The World's Easiest Portrait Enhancement Software. *Anthropics.com*. https://www.anthropics.com/portraitpro/photo_editing_software [2022. 09. 30.]

W2 = FaceApp 2022: Most Popular Selfie Editor. *Faceapp.com*. https://www.faceapp.com [2022. 09. 30.]

W3 = Beauty.AI 2022: Welcome to the First International Beauty Contest Judged by Artificial Intelligence Beauty.AI. *Beauty.AI*. https://beauty.ai [2022. 09. 30.]

W4 = This person does not exist. https://thispersondoesnotexist.com/ [2023. 02. 03.]

Watson, Imogen 2021: ‚Reverse Selfie': Dove's mission to combat social media's negative effects. *The Drum.com*. https://www.thedrum.com/news/2021/04/28/reverse-selfie-dove-s-mission-combat-social-media-s-negative-effects [2022. 09. 30.]

PSZICHOLÓGIA-PEDAGÓGIA

TARI ANNAMÁRIA

Manipulált képek és videók lélektana – a valóság kitágítása, avagy az illúziók valóságba emelése?

A 21. század társadalmi nárcizmusa és az online tér erőteljes jelenléte változásokat idéz elő a személyiség érzelmi működésében. A 21. századi jellemző viselkedés szoros összefüggést mutat az információs kor által nyújtott digitális gyorsasággal, az azonnali érzelmi szükségletkielégítés igényével, az érzelmi inkontinenciával és az ennek következtében megjelenő felnőttkori, mégis infantilisnak értékelhető indulatkezeléssel. Látjuk már, hogy az emberi érzelmeket nem hagyja érintetlenül az információtechnológia és a gyorsulás megatrendje sem. Fontos tehát, hogy mennyire engedjük el azt a gyeplőt, amely az érzelmi működésünk kontrollját jelentené. Most kell megtanulnunk, hogyan kell óvatosan és okosan élni ennek az új világnak az eszközeivel, hogy ne emberi kapcsolataink érzelmi minőségén csapódjon le egy eddig ismeretlen hatás. A mesterséges intelligencia, a deepfake olyan korszakot nyit, amelyben fontos lesz a saját realitásunk megtartása.

Kulcsszavak: online érzelmek, nárcizmus, realitás, érzékelés

1. BEVEZETÉS

Több mint egy évtizede áll a pszichológiai kutatások középpontjában az információs kor által megindított személyiségvonások változása. Ezek legfőbb vonulata a társadalmi nárcizmus elméletének megalkotása, a személyiségfejlődésre gyakorolt hatásainak elemzése és a lehetséges prognózisok felállítása. Ebben a fejezetben kísérletet teszek arra, hogy pszichológiai szempontokat adjak ahhoz a jelenséghez, amely a jövőnket – úgy tűnik – részben vagy egészben meghatározza majd: a valóság és a manipulált/kevert valóság érzékeléséhez és érzelmi hatásaihoz.

Néhány éve kutatják csupán a deepfake hatásmechanizmusát, az adatok nem mindig egyértelműek és megnyugtatóak. Érdemes tehát elgondolkodnunk azon a folyamaton, amelyben az érzelmek, a kapcsolati kommunikáció alakulása zajlik, követve az internet, a közösségi média és a tanuló algoritmusok, a mesterséges intelligencia fejlődését.

Az interneten elérhető, eligazító szándékú, *A virtuális marketing korszaka* című cikk azt taglalja, hogy „egy virtuális influencer sokkal rugalmasabb lehet egy együttműködés során, az ő kontentjét [...] azokra az igényekre alkotják meg, amit az ügyfél kér. Emellett pedig kevesebb az esélye annak, hogy a későbbiekben bármilyen [...] nyilvános botrányba keveredjen, ami kihat rá és az általa reklámozott márkákra" (Influencer Kisokos 2020). Valójában tehát a virtuális alkalmazottak részvétele a fogyasztói társadalomban már inkább tűnik hasznosnak, semmint különlegesnek, és a jövőre nézve is biztonságosabbnak látszik, hiszen nem fog kínos szituációkban esetlen mondatokkal védekezni. Mert nem kerül ilyenekbe. Ez a szolgáltatóipar praktikus és funkcionális fejlődése, amely nyilván a profitrealizálást és a biztonságos gazdasági környezetet preferálja.

„A csak virtuálisan létező Rozy 2021-ben várhatóan 1 millió dollárt keres a Sidus Studio X-nek, pedig mindösszesen 66 ezer követője van Instagramon" (W1). Baj van ezzel? Azt hiszem, erre még nincs jó és egyértelmű válaszunk. Egyfelől érthető az információtechnológia és a mesterséges intelligencia térnyerése, más kérdés viszont, hogyan fog az emberi psziché alkalmazkodni, ha lassan nehéz lesz eldönteni, akit lát, az valódi személy-e vagy sem, és az illető kommunikációjának tartalma igaz vagy sem.

Hogyan alakul majd a valóságérzékelésünk? De főleg, mi lesz a most felnövekvő nemzedékekkel? Leteszünk-e arról, hogy a realitásfunkciónk minden helyzetben vonalvezető legyen, és átadjuk magunkat a „mindegy is, ez a 21. század!" passzivitásának? Vagy szorongásaink támadnak, mert azt érezzük, kicsúszik a szilárd talaj a lábunk alól? És mindezeken túllépve, hogyan viselkedünk majd, ha olyan fejlesztésekhez jut az átlagos felhasználó, amelyek lehetővé teszik neki, hogy meghamisítsa mások létezését oly módon, hogy a deepfake segítségével átírja a róla kialakult képet?

Eligazodni az új világban nem lesz egyszerű. Éppen arra a képességünkre lenne majd szükségünk, ami alatt rángatják a szőnyeget: a józan észre, amely eddig legalábbis az emberiség fejlődésében segített abban, hogy álom és valóság, fantázia és realitás között legyen határvonal.

2. A DEEPFAKE ÉRZELMI HATÁSAIRÓL

Ahhoz, hogy a mesterséges intelligencia személyiségre gyakorolt érzelmi hatásain gondolkodjunk, nem árt visszatekinteni olyan kutatókra, akik már évtizedekkel ezelőtt kongattak vészharangokat azzal kapcsolatban: mit tud kezdeni az ember a körülötte zajló, egyre inkább felfoghatatlannak látszó technológiai fejlesztésekkel.

Marshall McLuhan szerint: „A környezet megváltoztatásával a média egyedi arányokat létesít érzékeink között. Bármely érzékünk kiterjesztése megváltoztatja gondolkodásunk és cselekedeteink módját – a világ észlelésének módját. Ha ezek

az arányok változnak, az emberek is változnak" (McLuhan–Fiore 2012: 41). Ezek szerint minden médium – amit ő az emberi érzékszervek kiterjesztésének tekintett – megváltoztatja kapcsolatunkat a minket körülvevő világhoz, „az emberek bevonódnak egymás életébe, közvetlenül és szakadatlanul ömlik ránk az információ, melyet alighogy feldolgoztunk, máris követ az új, aztán még újabb adag" (McLuhan–Fiore 2012: 63).

Ahogy Fodor és L. Varga kifejtik, „McLuhan médiaelméletének [...] hatástörténeti jelentősége voltaképpen abban áll, hogy a kilencvenes évektől kezdve mind hangsúlyosabbá és nyomasztóbbá válik az a felismerés, hogy nem a médiumok vannak bekötve az emberbe, hanem az ember a médiumokba, ily módon a médiumoknak az ember már nem a szubjektuma. Így tekintve a technikai újítások kizárólag egymásra vonatkoznak, illetve kizárólag egymásra adott válaszok, és az ember individuális vagy kollektív testétől teljes mértékben leválva zajló fejlődésnek az eredménye az érzékekre és szervekre gyakorolt elemi erejű hatás" (Fodor–L. Varga 2018).

Purnell (2020) ezt azzal egészíti ki, hogy Guy Debord azt állította, hogy a látványosság társadalmában élünk; Michel Foucault kifejtette, hogy a modern életet hogyan hatja át a megfigyelés; a pszichológus Jacques Lacan bevezette a „tekintetet", hogy megmagyarázza a szorongást, amit az a felismerés okoz, hogy mások számára láthatóak vagyunk (Purnell 2020).

Annak magyarázata, hogy miért vezérel bennünket a „hiszem, ha látom" attitűd, érthetővé válik, ha belegondolunk, hogy a kommunikáció fejlődésében a szájhagyományt váltotta fel a nyomtatott szöveg. „Az igazi választóvonal azonban a tizennyolcadik század volt, amikor az írástudás fellendült, feltalálták az újságnak nevezett újszerű médiumot, és a levelek szabadon terjedtek. A modern élet egyre inkább »szemközpontúvá« vált az olyan új médiumok megjelenésével, mint a fényképezés, a film, a televízió és az internet" (Purnell 2020, a szerző fordítása).

Ha tehát így gondolunk a deepfake jelenségére, mint olyan eszközhatásra, amely a legfőbb vizuális realitástámaszunkat teszi próbára, akkor különösen fontos lesz, hogy milyen képességekkel kell rendelkeznünk ahhoz, hogy eldönthessük: igaz vagy hamis-e, amit látunk. Így csalatkozhatunk abban, amit látunk, a médiában, a realitásban, és ez elvezethet ahhoz, hogy az igazságérzetünkben sem bízhatunk igazán. Különösen azért, mert ezek előállítása és ezért megjelenése is egyre könnyebb, gyakoribb: „A deepfake – az audiovizuális tartalmak hiperrealisztikus utánzata – előállítása egyre egyszerűbbé és olcsóbbá válik, ez nyilván egyet jelent azzal, hogy számuk az egekbe szökhet. Mindössze egy év alatt, 2019 és 2020 között a deepfake-ek száma 14 678-ról 100 millió videóra nőtt, ami 6820-szoros növekedést jelent" (Köbis–Starke–Soraperra 2021, a szerző fordítása).

Megtéveszthet saját meggyőződésünk is, ha arra gondolunk, hogy „soha nem dőlnék be egy deepfake-nek". Ez olyasfajta magabiztos attitűd, amelyet valójában

talán nem is vetünk alá realitásvizsgálatnak. Érzelmeink világában így tudunk előállítani annyi önbizalmat, amelytől azt reméljük, megtart minket a valóságban.

A Center for Humans and Machines (Max-Planck-Institute for Human Development) és a CREED (University of Amsterdam) kutatói által közölt tanulmányban (Köbis–Doležalová–Soraperra 2021) a válaszadók túlnyomó többsége komolyan bízott a deepfake felismerésének képességében. Ez a magabiztossági szint azonban jelentősen meghaladta a valós eredményeiket. Amikor az emberek saját percepciós képességeikre támaszkodtak (hallás, látás), már nem tudták megbízhatóan felismerni a deepfake-et. Még akkor is túlzottan magabiztosak voltak, amikor anyagilag ösztönözték őket arra, hogy minél pontosabban megjósolják a saját felismerési képességeiket. Ezek az eredmények az emberek túlzott magabiztosságát mutatják, amiben nyilván jelen van az önbizalom és az érzelmi biztonság attitűdje, vagyis azé a meggyőződésé, hogy „senki nem vihet be az erdőbe". De sajnos az is kiderült, hogy a deepfake negatív következményeinek tudatosítása nem növeli a résztvevők észlelési pontosságát ahhoz a kontrollcsoporthoz képest, amelynek tudatosságát nem emelték. Ez azt sugallja, hogy a deepfake felismerése nem annyira érzelmi motiváció, mint inkább képesség kérdése lesz (Köbis–Starke–Soraperra 2021). A kérdés tehát az, vajon milyen képességünket kell majd fejlesztenünk a realitásunk megőrzéséhez.

Aronson nagyon szellemesen fogalmazza meg, hogy: „Az emberi elmében eredendően optikai és pszichológiai vakfoltok vannak, és az elme egyik legügyesebb trükkjeként azt a kellemes téveszmét táplálta belénk, hogy nekünk személy szerint nincs egyetlenegy sem" (Aronson–Tavris 2009: 44).

Ha az a meggyőződésünk, hogy a szemünknek és a döntéseinknek hihetünk, akkor meglepő eredménynek tűnhet egy friss ausztrál kutatás (Moshel et al. 2022), amely szerint agyunk (tudatunkkal ellentétben) nagyon is felismeri a valótlanságot. A kutatók a vizsgálatokhoz EEG-t használtak, amely lehetővé teszi valós időben az idegsejtek elektromos aktivitásának regisztrálását, azért, hogy a központi idegrendszer működését figyelve, feltérképezhetőek legyenek a kísérleti személyek percepciói. Eredményeik szerint az emberi agy képes felismerni a mesterséges intelligencia által generált hamis arcokat, akkor is, ha az ember nem tudja megmondani, melyik arc valódi és melyik hamis. A hamisítványokat a résztvevők agyi aktivitása alapján az esetek 54 százalékában sikerült azonosítani. Amikor azonban arra kérték a résztvevőket, hogy szóban is azonosítsák ezeket, csak az esetek 37 százalékában tudták ezt megtenni. A kutatás konklúziója szerint az agy (ha nem is tudatosan, de) felismeri a különbséget a deepfake és a hiteles képek között (Moshel et al. 2022). A kutatók két kísérletet végeztek: az egyikben a résztvevőknek 50 képet mutattak valódi és számítógép által generált hamis arcokról. Arra kérték az alanyokat, hogy azonosítsák, melyik a valódi és melyik a hamis arckép. A másik csoport viszont úgy nézte végig a képeket, hogy nem tudtak az igaz/hamis jellemzőkről. A kutatók ezután összehasonlították a két kísérlet eredményeit,

és azt találták, hogy az emberek agya jobban azonosította a nem valódi arcot a valósakhoz képest. Ez alapvetően reményt adó hír, még akkor is, ha további kutatásokat igényel, mert ezek szerint támaszkodhatnánk az agyunkra, ha képesek lennénk a meggyőződésünk helyett az ösztönös megérzéseinkre hagyatkozni. Ez utóbbi képességünk az élet sok területén megmutatkozik.

Gondoljunk olyan döntéseinkre, amelyeknél a fejünkben megszólal egy „kis csengő", és arra figyelmeztet, hogy valamilyen „zavarjel" miatt nem látjuk feltétlenül jól a helyzetet. Ám az esetek többségében az emberek ezt az intuitív szignált figyelmen kívül hagyják, és elterelődnek más érzelmi szempontok mentén (például az önigazolással, aminek eredményeképpen létrejönnek a „szerintem valódi!" megállapítások). Ennek magyarázataként gondoljunk arra, amit Aronson és Tavris mond a gondolkodás torzulásáról, „ami már azzal megkezdődik, ahogyan az agy feldolgozza az információt. [...] Az emberek logikus gondolkodásáért felelős agyterülete a disszonáns információk hatására gyakorlatilag kikapcsolódik, míg az agy érzelmi területe boldogan aktiválódik, amikor a konszonancia helyreáll. Ezek a mechanizmusok szolgáltatják a neurológiai alapot a megfigyeléshez, mely szerint elhatározásunkat nehéz megváltoztatni, mihelyt eldöntöttünk valamit" (Aronson–Tavris 2009: 26–27).

A deepfake térnyerése során azonban belső iránytűnkre – ezek szerint – jobban szükségünk lesz, mint valaha. Ha fejlődni szeretnénk, disszonanciáink tudatosabb észlelésével és véleményünk alaposabb átgondolásával kell kezdenünk.

3. A GYEREKEK ÉRZELMEI ÉS KOGNITÍV JELLEMZŐIK

Több éve foglalkoztat az információtechnológia személyiségre ható vonásainak értelmezése, mert a felnőttekhez képest a legfontosabb kérdés mégis az, hogy a legfiatalabbak generációja és a még meg sem születettek hogyan tájékozódnak majd a világban, ha az érzelmi biztonsághoz, a realitásfunkció alapjául szolgáló percepcióikhoz már kapcsolódik az információtechnológia.

Egy 2022-es ismeretterjesztő cikk (Varga 2022) utal arra, hogy a hangalapú aszszisztensek, köztük a Google Home, az Amazon Alexa és az Apple Siri gyors térhódítása a kutatók szerint hosszú távon káros hatással lehet a gyermekek szociális és kognitív fejlődésére, különösen az empátiára, az együttérzésre és a kritikai gondolkodásra. A gyerekek a hangvezérelt okoseszközöknek emberi tulajdonságokat és viselkedést tulajdonítanak, ami a társas viselkedésük fejlődése során rögzíthet nem megfelelő válaszokat, mert nem érzelmi alapú emberi kommunikációt hallanak, de akként azonosítják.

Konok Veronikáék (ELTE Alfa Generáció Laborban végzett) vizsgálata is ráerősít arra, hogy hosszú távú hatása lehet az érintőképernyő és hangalapú alkalmazások használatának. Összehasonlítva érintőképernyős eszközöket gyakran hasz-

náló és nem használó óvodás korú gyerekek szelektív és megosztott figyelmi és szociokognitív képességeit, eltérő eredményeket kaptak aszerint, hogy a gyerekek a vizsgálat előtt digitális játékkal vagy nem digitálissal játszhattak. Eredményeik azt mutatták, hogy az érintőképernyős eszközök rövid és hosszú távú használata és azon belül a digitális játékokkal való játék lokális figyelmi fókuszhoz vezet, vélhetőleg azért, mert a digitális képernyők lokális információkban gazdagok, és ritkán látszódik egyben az egész kép. A gyors digitális játék azáltal, hogy párhuzamosan több ingerre kell figyelni, a megosztott figyelmet fejleszti, de a szelektív figyelmet nem, amely abban segít, hogy az információk közül képesek legyenek a lényegesre fókuszálni. További fontos megállapításuk, hogy az érintőképernyős eszközök használata elveszi az időt – az ebben az életkorban szükséges – társas tevékenységektől, ami magyarázhatja, hogy nehézségeik lesznek a komplexebb szociokognitív képességek terén, például helyzetfelismerés, helyzetek kontextusának az értelmezése (Konok et al. 2020).

Ha feltesszük a kérdést, hogy miért fontosak ezek az eredmények, akkor gondoljunk arra, hogy a teljes kép, vagyis a globális fókusz abban segít, hogy a világot ne összefüggéstelen pontok halmazaként érzékeljük, hanem rögtön észrevegyük benne az értelmes alakzatokat. A hangalapú, érintőképernyős eszközökön játszó gyerekek nehézségekkel találhatják szembe magukat, amikor a világ történéseinek, az online tér által nyújtott virtuális valóság képeinek megértésére lenne szükségük. Nem is beszélve arról, hogy már eleve egy olyan kevert világban nőnek fel, ahol a valóságos és nem valóságos, emberi és nem emberi szétválasztásának képességéhez éppen a realitás, vagyis az offline valóság nyugodt megismerése, explorációja, a kommunikáció és az érzelmek fejlődése lenne szükséges. Azon is elgondolkodhatunk, hogy vajon a kevert valóságban létezés megítélése az offline gyerekkorral jellemezhető idősebb generációk problémája-e, mert ők még jobban ragaszkodnak az igazságérzetükhöz?

4. AZ ONLINE TÉR HATÁSA AZ ÉNFUNKCIÓKRA

Amikor az okoseszközök használata egyre erőteljesebbé vált, a klinikai pszichológus kutatókat egyre inkább kezdte foglalkoztatni, hogy mi és hogyan változik majd ennek következtében. Az online tér közösségimédia-fejlesztései, a tanuló algoritmusok folyamatos jelenléte komoly tényező ma már abban, hogyan alakulnak az emberi kapcsolatok, hogyan változik a kommunikáció, és abban, hogy az érzelmek változása ebben a párhuzamos valóságban miként formálódik ahhoz képest, hogy az illető mit gondol saját (reális) személyiségjellemzőiről.

Az azonnali érzelmi szükségletkielégítést lehetővé tevő okostelefonok – a gyorsasággal, az érzelmi inkontinenciával, a multitasking működéssel, továbbá az intimitáshatárok eltolódásával – jelentős változásokat indíthatnak el a fiatal generá-

ciók személyiségfejlődése folyamán, és indítottak már el a felnőtt személyiségek életében is. Gondoljunk a troll jelenségre, amelyben látszik a lélektanilag infantilis (azonnali) düh és indulat, mert a posztok és kommentek az impulzuskontroll oldódásának eredményeképpen jönnek létre.

Előadásaimon sokszor beszélek a deepweb, a bosszúpornó, az online bullying és karaktergyilkosság jelenségéről, amelyekhez a technológia nyújt segítséget, és általában az átlagfelhasználók megrökönyödését tapasztalom, akik szörnyülködve kérdezik: „ilyenek hogyan jöhetnek létre?" A válasz néha egyszerűen az, mintha a fejlesztő programozók csak végeznék a dolgukat, viszont újabb és újabb alkotásuk nagyszerűsége talán elhomályosítja a lényeget: hogy milyen társadalmi, érzelmi hatást fognak gyakorolni az emberek életére.

Amikor a deepfake, mint lehetőség, elérhetővé válik, nem lesz meglepő, ha emberek milliói kapnak valójában olyan lehetőséget, amelyben az indulataikat már nemcsak a bullying, hanem kifinomultabb és a valóságot elég hűen másoló mesterséges intelligencia fogja támogatni. A gyorsulás megatrendjében fontos kérdés lesz, hogy az érzelmi fékek és ellensúlyok hogyan működnek majd, mert a lassúság az élmények, érzelmek átélése és feldolgozása folyamán igen fontos jellemző. Érzelmeket feldolgozni, élményeket szerezni az internet előtti korszakban offline helyzetekben lehetett, amelyekben a másik emberrel való kapcsolat a realitásban történt, az érzelmek megélésének lehetőségével. Ma az online térben az élmények sokasága történhet szinte egy időben, ami nélkülözni fogja a feldolgozáshoz szükséges nyugalmat és időt. Az érzelmi elárasztódás így inkább telítettséghez vezethetnek, ami egyet jelenthet azzal, hogy abból építkezni szinte lehetetlen, vagyis a személyiség érzelmi kapacitásai hamarabb kimerülnek, mintsem töltődnének.

Az énfunkciók (éntudat, énkép: testkép, szociális kép, önbecsülés, identitás, az én-ideál és az önkontroll – önértékelés) lassú fejlődése, a realitással való összevetése, a szociális közegben megtapasztalt érzelmi élményekkel való fejlődése, ha nagyrészt az online tér identifikációs közegében zajlik, akkor következményképp változást fog mutatni a régi offline metódushoz képest.

Az online tér döntően vizuális szcéna, amelyben értelemszerűen a képekben és képekkel kommunikálás, azok folyamatos előállítása, a korábbi verbális szcéna lassúsága helyett a gyors váltakozások területére viszi a személyiség megméretését. A kutatások nagy részében (Course-Choi–Hammond 2021; Vogel et al. 2014; Timeo et al. 2020) már látszik, hogy ez a mértékű vizualitás nincs feltétlenül jó hatással az énkép és identitás alakulására, mert a folyamatos énprezentáció (a közösségi média profiljai) valójában az énideál mutációit tarthatják fenn, ami az érzelmi szükségletkielégítés azonnaliságának igényével együtt infantilizálhat, ami nehezítheti a realitás pozitív definiálását és megélését. Az irrealitás és a realitás viszonya tehát elcsúszhat, miközben az érzelmi stabilitáshoz a realitás szükséges.

Gondoljunk bele, hogy az első szinten (vizualitás) is csúszkáló realitás még további kihívásokkal terhelődik, például a deepfake esetében. Mennyire tolódhat el

a valóságérzékelésünk, ha már a szemünknek és a fülünknek se hihetünk, és ez vajon megnöveli-e annak a narcisztikus attitűdnek a megjelenését, amely az önbizalmat táplálja egy irányú nélküli életben? És hogyan alakul majd így a realitásba vetett bizalmunk, ami pedig az élet egyik tartópillére? Változik-e a szeparációs szorongás mértéke, annak függvényében, hogy a világgal, igazságokkal, valósággal kapcsolatos bizonytalanságok is terhelik majd?

A közösségi média felületein az információk láthatósága fokozhatja az önbecsülést, különösen akkor, amikor valaki saját maga szerkeszti az énjéről szóló információkat, ami arra vall, hogy a digitális önmegjelenítés módosíthatja az önértékelést. Ez az állandó transzparencia hozza magával az állandó kapcsolatban levést, a rövid (és a netspeak miatt egyre rövidebb) információközléseket (posztok), illetve azt az érzelmi igényt, hogy az „olvasók" azonnal reagáljanak. Ez a „lájkgyűjtés" olyan eleme az online térnek, ami egyértelműen a szeparációs szorongás kiküszöbölése érdekében történik. A bármikor megkapható „pozitív válasz" elvárása erős, és adott esetben az illető aktuális érzelmi állapotára jelentős hatással van. Ezt ismerjük „narcisztikus buborékként", amely azon ismerőseink csoportja, akikkel egy véleményen, érzelmi álláspontón vagyunk, és amely a buborék természetéből következően csak nehezen enged be más, eltérő (esetleg a realitást jobban képviselő) véleményt.

Látjuk tehát, hogy a közösségi média a „sosem kell egyedül lennem" érzelmi attitűdjét erősíti, miközben egy személyiség jó érzelmi működésének alapvető vonása, hogy képes az egyedüllét elviselésére, a saját gondolataival való együttlétre anélkül, hogy ehhez külső segítséget hívna. A szeparáció tűrése egyben alapköve lesz az emberi kapcsolatok stabil építésének, a partneri viszonyok optimális kialakításának, mert nem igényel a másik féltől folytonos visszaigazolást. Ez egyben azt is jelenti, hogy stabillá teszi az önértékelést, mert a belső érzelmi egyensúlyt képes egyedül is fenntartani, külső támogatás nélkül. De ma az online tér csoportnormája igen erős és igen vonzó. Azt a szorongásmentesítést kínálja, ami sok esetben nem segít a realitás megismerésében és a belső biztonság fenntartásában.

Az önértékelés alakulása szempontjából a társas kapcsolatok kiemelten fontosak. A szociális visszajelzések régen sokat segítettek abban, hogy egy-egy külső információt, véleményt, értékelést beemelve fejlődjön valaki. Ma komoly veszély azonban, hogy a netspeak igazán nem alkalmas érzelmi közlésekre, a „puszifejek és szívecskék" valójában nem rendelkeznek tartalommal, inkább csak jeleknek tekinthetők. És miközben a lájkok tengerében él mindenki, a valóságban jóval inkább a társadalmi nárcizmus, a rivalizáció és a nem kevés agresszió van jelen.

5. A TÖMEGMÉDIA ÉS AZ ASSZERTÍV ÉRZELMEK ALAKULÁSA

A televízió és az internet is a tömegmédia egy-egy eleme, amelynek egyik fő jellemzője, hogy sémákat mutat, amelyekben az egyén fellelheti magát, és amit bemutat, az mindenkit érint. Így válik lehetővé az azonosulás kultusztárgyakkal, mozgalmakkal, trendekkel, amelyeket a tömegmédia forgatókönyvei diktálnak, tehát az egyén külső szemlélő és belső (azonosuló) résztvevő is lehet. A kortárs kultúra tehát hatással van a kapcsolati sémákra, emberi viszonyokra és életmódra. Csak példaként nézzük meg, hogy a kiskamasz korosztály most olyan kortárs kultúrát lát maga előtt, amelyben intézményesül az érdek kielégítése – mindenekfelett. A valóságshow-k (Guld 2022) népszerűsége töretlen, annak ellenére, hogy sokszor a szakértőkkel egyetemben a nézőik is tartalmatlannak és ártalmasnak ítélik ezeket a műsorokat. Lélektani hatásaik közül kiemeljük azonban, hogy a kiskamasz korosztály számára e műsortípus szigorú és kérlelhetetlen kiközösítései, a szereplők agressziója azt példázzák, hogy „jó keménynek kell lenni, ha akarsz valamit". Ez alapvetően még nem lenne baj, hiszen ez a kitartás és a feladattartás együttesének értelmét is szolgálhatná. A probléma inkább ott van, hogy ezt bármikor mások kárára is meg lehet tenni. Sőt ez a kívánatos viselkedés, ha valaki nyerni akar. Ez viszont már olyan forgatókönyv, amely arra nevel, hogy a személy a saját érdekeit sose tévessze szem elől. A társas viselkedésben viszont az asszertív magatartás az ajánlott, különösen konfliktusmegoldás és problémakezelés esetén. A nem mások kárára megvalósított célok „érzelmi útvonalai" sokkal harmonikusabban illeszkednek be a csoportlétbe, mint az agresszív forma, amely előbbutóbb benyújtja a számlát, és magányra ítélhet.

A mesterséges intelligencia fejlesztései, különösen a deepfake sokak kezében válhat olyan fenyegető eszközzé, amely nem az asszertív, hanem az agresszív viselkedés felé tolódik, mert – ahogy a bosszúpornó, az online bullying esetében is – az áldozattal szembeni érzelmek feloldódnak abban a narcisztikus haragban, impulzuskontroll-problémában, ahol az elkövető feljogosítva érzi magát egy másik ember lejáratására, bántására, lealázására, adott esetben tönkretételére.

Nem akarok szörnyű disztópiát festeni, csak a lehetséges társadalmi diagnózisok között jelenleg nem sok remény van arra, hogy soha senki nem fogja negatív célokra használni a deepfake lehetőségét. Minél bizonytalanabb a külső világ, annál inkább megnő az agresszió veszélye, mert a növekvő szorongások, ha nincs kanalizációs lehetőségük, akkor eszkalálódnak. Ehhez pedig a mesterséges intelligencia eszközöket nyújthat.

Belátható, hogy a közösségi média felületein mindenki saját „feljavított" énjét mutatja, és ez a kép áll kapcsolatban mások hasonlóan a reprezentációra szánt képeivel. Mindez írásos és vizuális közlés, amelyben nagyon fontos a közönség megléte, és ami összefüggésben áll az illető önbecsülésével, aktuális érzelmi állapotával. De ezek a közlések, bár döntően pozitívak, nem helyettesíthetik a valódi *(face-to-face)* hely-

zetek fontosságát, bármennyire is szeretnénk. Az online tér kapcsolati rendszerében megtanult szociális ismeretek azonban nem feltétlenül segítenek az offline helyzetek jó kezelésében. Például Jean M. Twenge (2018) kutatási adatai szerint az Y és Z generáció gyengébbnek mutatkozik az interperszonális képességek terén.

Valójában a jó személyes hatékonyság (asszertivitás) biztosítja majd a környezettől megkapott elismeréseket, amelyek mint jó érzelmi tapasztalatok adják majd a felnőttkori sikeresség egyik fontos alapvonását, azt az érzést, ami arról szól „meg tudom csinálni". Azok, akik sikeresek, rendelkeznek kellő bátorsággal és erőfeszítésekre való képességgel, ami számukra elérhetővé tesz reális célokat. Az asszertív emberek törekszenek arra, hogy felismerjék a kitűzött céljaik elérésének kritikus sikertényezőit, ami azt jelenti, hogy kiválasztják a jelen működési modellből, mely mechanizmusok azok, amelyek alapul szolgálhatnak a cél eléréséhez, illetve melyek azok, amelyek akadályozzák őket, és amelyeknek meg kell változniuk a siker érdekében. Azonban nem ártanak másoknak, nem mások vállán kapaszkodnak a csúcsok felé, hanem folyamatosan értékelik és összevetik tapasztalataikat az eredményekkel, majd döntést hoznak egy új irányvonalról (Tari 2013).

Az asszertivitás nem feltétlenül az online tér jellemzője, és különösen nem a közösségi média felületeié, ahol az énprezentációk találkozása adja a kapcsolat alapját és ahol az éretlenebb személyiségek (gyengébb önkontrollal) valójában fésületlenül ontják a válogatás nélküli érzelmeiket.

6. HOGYAN ALAKULNAK AZ EMBERI KAPCSOLATOK ÉRZELMEI AZ ONLINE TÉRBEN?

Ma egy újonnan megismert embert mindenki megpróbál megtalálni a neten. Megnézik, milyen az oldala, milyen képei vannak fenn, kikkel barátkozik, mit szeret, milyen zenét hallgat stb. Ez olyan információtömeg, amihez nem kell az illető engedélye, hiszen ez mind szabadon elérhető tartalom, gyakran valóban az, ha az illető elfelejtette a „nem publikus" funkciógombot használni.

Így a valakiről kialakult első benyomásunk nem „múlt századi", hanem az információs korra jellemző. A megszerezhető információk nagy részénél tehát az aktivitást nem a beszélgetés és a kérdezősködés jelenti, ami egyben a kölcsönösséget is biztosítja, hanem egy egyoldalú folyamat, a „keresés" lép ennek a helyébe. Úgy vélem, ez befolyásolja az „első benyomást" is, ami ilyen módon nem feltétlenül a realitásban alakul, hanem megelőzi egy online kép. Amikor egy idegennek nézem az oldalát, akkor kialakítok róla valamilyen képet. Nyilván ebben szerepe lesz a megelőző tapasztalataimnak, a nézeteimnek, a véleményemnek s minden olyan tudásnak, amit a „hívószavak" elindítanak.

Solomon Asch híres kísérletében (1946) arra kérte a kísérleti alanyait, hogy jellemezzenek egy ismeretlen személyt aszerint, milyennek gondolják egy tulajdon-

ságlista alapján. Az egyik csoport listáján az intelligens, szorgalmas, impulzív, kritikus, makacs és irigy sorrend volt látható, a másikén viszont éppen fordítva: irigy, makacs, kritikus, impulzív, szorgalmas és intelligens. Ezután megkérte mindkét csoportot, hogy értékelje a személyt aszerint, mennyire képzeli boldognak, társasági lénynek. Akik a kedvező tulajdonságokat kapták a lista elején, jóval pozitívabban ítélték meg az illetőt, mint azok, akik a negatív kezdést láthatták. A kedvezőtlen ítéletet tehát komolyan befolyásolták a korábbi tényezők (benyomások), a realitástól függetlenül. Asch ezt a jelenséget „elsőbbségi hibának" *(primacy error)* nevezte el. Mi történik valójában? A kísérleti személyek a tulajdonságlista alapján már elkezdtek összerakni egy képet arról a személyről, akit nem ismertek. Az, hogy pozitív vagy negatív volt-e számukra a kezdő három tulajdonság, a továbbiakat is befolyásolta. A jelenséget valójában nagyon sokszor átéljük. Amikor olvasunk vagy hallunk valakiről, akit nem ismerünk, csak érkezik róla valamilyen „tulajdonság", azonnal képesek vagyunk arra, hogy egy személyiségképet alkossunk. A hiányzó információk és a realitás nem zavarnak minket, mindössze egy vágy vezérel, ez pedig az ítéletalkotás. Ez a jelenség összefügg a kommunikáció „megszaladásával" (Miklósi 2010: 164). Miklósi gondolatmenete szerint az embercsecsemő az a teremtmény, aki már születésétől kezdve „nagyon erős preferenciával rendelkezik a másokkal való kapcsolatteremtésre". A gyerek már a beszéd megjelenése előtt aktívan keresi a lehetőséget a kommunikatív kapcsolatra, elsősorban az anyával. A néhány órás csecsemő már utánoz, sokféle hang kiadására képes, és folyamatos a kapcsolati törekvése. A beszédhez azonban szükség volt a másik fizikai jelenlétére, ami a média megjelenésével megváltozott. Az utóbbi időkben – mivel korábban a médiát az információ tárolásának fejlődése jellemezte – azt mondhatjuk, hogy ma a fejlesztések a kommunikáció terjesztésének hatékonyságát érintik. Amit látunk: az emberben meglévő, evolúciósan szelektált és egyedfejlődésileg a csoport által megerősített kommunikációs preferenciát magával ragadja a kulturális-technikai fejlődés új eszköztára (Miklósi 2010: 165–166). A megnövekedett kommunikációs lehetőségeknek az ember nem tud ellenállni, és előáll a „megszaladás" jelensége, amelyben azonban nyilvánvaló, hogy a túl sok emberrel való kapcsolattartás elsekélyesíti a szociális viszonyokat. „Az elektronikus kapcsolatok »fizikai megerősítés« hiányában felszínessé válhatnak, és frusztrálhatják az embereket" (Miklósi 2010: 166).

Ez tehát az a folyamat, amely során a negatív érzelmek és feszültség kezelése – mint az erőfitogtatás és rivalizálás – a technológia támogatásával történik. Ha most elgondolkodunk a technológia ilyen jellegű „segítségén", akkor dilemmákba ütközhetünk. Egyfelől érthető, hogy az egyre agresszívabb világban az online tartalmak is nagyobb kiváltott érzelmi potenciálra utaznak. Másfelől viszont aggasztó, ha valaki úgy „használja el" az empátiáját, hogy más emberek őrültségeit vagy szenvedéseit mutató képsorokat néz. Ebben az esetben is érvényes a „katarzismodell", vagyis a látott képek bámulása közben átélt agresszió segít csökkenteni

a belső feszültségeket. Mindez azonban erősen függ a szociokulturális háttértől, az iskolázottságtól, a családi érzelmek és kommunikációs rendszer milyenségétől, valamint az egyéni karaktervonásoktól (Somlai 1997; Tóth 2011).

De a digitális csatornán beindított kommunikációnak lesz még egy sajátossága, ami a gyorsasággal függ össze. Az érzelmek sokkal hamarabb válnak intenzívvé, mint egy fizikai kapcsolatban. A monitor ugyanis olyan felületté válik, ami teret ad a fantáziának, az ábrándozásnak, a tudattalan érzelmek korlátlan megjelenésének. Nincs jelen a másik ember, csak a képe. Minden vele való érzelmi kapcsolat tehát olyan lélektani térben megy végbe, amelyben a képzelet játssza a főszerepet, és nem a valóság. Ennek a belső képnek a csiszolása, alakítgatása aztán törvényszerűen vezet oda, hogy felgyorsulnak az események, és olyan érzelmek is megjelennek, amelyek hagyományos szituációban jóval később születnének meg. És könnyebben válik az indulatok terepévé az az írásos kapcsolódás, amely sok esetben a személyesség deficitjével jellemezhető.

7. AZ ONLINE IDENTITÁSOKRÓL

„Új énjeink némelyike – mondja Wallace (2006) –, határozatlanul megformált, nagyon ideiglenes, éppen csak több próbaidentitásnál." Ezektől a különböző visszajelzések tükrében megszabadulhat a tulajdonos, mások akár gazdagon kidolgozott személyiséggé nőhetik ki magukat. Folyamatosan törekszünk arra, hogy másokban jó benyomást alakítsunk ki magunkról, de másokkal kapcsolatban is nagy sebességgel jutunk el következtetésekig a legkevesebb információ alapján, anélkül, hogy ellenőriztük volna feltételezésünk hitelességét. Hajlamosak vagyunk a másik személlyel kapcsolatos egyetlen pozitív vagy negatív információ alapján feltételezni, hogy a személy többi tulajdonsága is összefügg ezzel, mintegy burkolt személyiségelméletként, amely bizonyos személyiségvonások összefüggéseiről alkotott nézeteinket foglalják magukban (ha híres, akkor gazdag, ha gazdag, akkor kiegyensúlyozott, ha kiegyensúlyozott, akkor boldog). Egy olyan környezetben pedig, ahol ráadásul nem is kell felfedni valódi énünket, korlátlanul játszhatunk különféle énjeinkkel, mindenki megalkothatja saját jelmezét, jó esetben nem túl távol saját, valódi identitásától (Wallace 2006: 78).

Messzire jutottunk már ettől a felfedezéstől, amely identitáslaborként definiálta az internet – mindenki által elérhető – terét. Ez a felület olyan hellyé vált, ahol tényleg összetalálkoznak az emberek egymás online leképezéseivel. Akik ismerték egymást, akik sosem találkoztak volna, és akik sosem látták egymást eddig. De most látják. Olvassák egymás ilyen-olyan sorait, nézik az ilyen-olyan képeit. Valójában az egymást jól ismerő emberek is megláthatnak valamennyit egymás eddig rejtett vonásaiból. Hihetetlen mennyiségű az az érzelmi motiváció, amelyben az „én is leszek valaki", „megmutatom magam mindenkinek" a vezérelv, vagyis lep-

lezetlenül az exhibicionizmusról szól. Valóban ennyien akarják magukat viszontlátni egy-egy oldalon? A lehetőség mindenkinek adott. Nincs színvonalelvárás, korrekció, de vannak durva helyesírási hibák és közönséges mondatok. Vannak szép és szépelgő tartalmak, okos és okoskodó vélemények, megértés és izzó gyűlölet. Mindez ott van abban a világban, ami új szabályok szerint működik.

Lehet, hogy régen a kultúra a magas elefántcsonttorony lakója volt, és most leköltözött a földszintre, és beleolvadt a hosszú farok jelenségébe (ez a kis menynyiségben eladható, de nagyon nagy sokféleségben rendelkezésre álló termékek gyűjtőfogalma). Látszólag tényleg megközelíthetővé vált szinte minden és mindenki, és mindenből rengeteg van.

Kevés olyan híres ember, sportoló, művész, tudós van, akinek ne lenne valamilyen online felülete. Bárki írhat nekik, és még akkor is meglesz a megelégedettség élménye, ha tudtán kívül egy olyan hivatalos oldalra írt, ahol majd egy robot vagy egy admin válaszol. Ha a „megtehetem" élménye kapcsolódik az „én is vagyok olyan ember, mint ő" leértékelő attitűdjével, akkor megjelennek azok a levelek és kommentek, posztok és üzenetek, amelyekben az „átlagember", aki nem rendelkezik kimagasló képességekkel, aztán megmondja a magáét. És ebben a látszatdemokráciában olybá tűnik: egyébként miért is ne mondhatná meg? Tényleg joga van hozzá. Mindenkinek joga van ahhoz, hogy legyen véleménye. Eddig is volt, csak éppen nem láttuk. Társadalmi méretekben ennyire nem látták egymást az emberek, mint most, amikor a bármilyen stílusban monitorra vetett sorokat és képeket nézegethetjük. Tudtuk, hogy az emberek nem egyformák, csak ennyire nem látszott, hogy milyen az, amikor a sokféle képességű ember egy rugóra mozog, márpedig ez a narcisztikus énmegjelenítés, a láttatás és a rivalizáció motivációja.

Andrew Keen (2007) – az egyik nagy kritikus – úgy véli, hogy a közösségi média valójában olyan narcisztikus élményhez juttatja tagjait, amelyben már nem az írástudó hivatásosok sorait olvassák a tömegek, hanem fordítva, mert ma mindenki lehet gyártó és fogyasztó egyben. Bárki írhat könyvet, cikket, kritikát, blogot, tehát minden további nélkül válhat íróvá, producerré, újságíróvá, kritikussá és szakértővé, filmes szakemberré, egészségügyi tanácsadóvá, nem is beszélve az életvezetési tanácsadók egyre népesebb táboráról. Ahogy ő fogalmaz: megjelent a „szent amatőr", tehát nincs szükség a profikra, cenzúrára és ezzel együtt az önreflexióra vagy önkritikára sem (Keen 2007). Ez volt talán az a mérföldkő, ahol a narcisztikus érzelemvezéreltség az internet elfogadott jellemzőjévé vált, és amely mára kitermelte az online bullying, a trollok jelenségét, és felveti azt a komoly kérdést, milyen világ lesz az, ahol a valóságot már nehéz lesz felismerni.

Ma – a közösségi életben – a közösségi vált kultúrává. Segít abban, hogy ne érezzük át a magányt, miközben ezt az emberi érzést meg kellene ismernünk, meg kellene tanulnunk, és nem kellene rettegni tőle. A narcisztikus működés mindenkit felhajt az online térbe, ahol a jel/zajrengetegben folyamatosan meghatározhatja

önmagát, átélheti, hogy „együtt van". Mert közben másokat is láthat, mindenki mást, aki szintén ott bolyong abban a világban.

A legnagyobb veszély, amit Umberto Eco (2008) állít, hogy olyan hipervalóságban élünk, amelyben a média reprezentációi fontosabbá válhatnak a tényleges valóságnál. És miután könnyebb a hozzájutás, ezen illúziók megszerzése és átélése egyre vonzóbbá válhat.

8. AGRESSZÍV A VILÁG IS, AMELYBEN ÉLÜNK

A 21. század jellemző tulajdonságai: a bizalmatlanság, a közöny tulajdonképpen azt mutatják, hogy az empátiát az ember fenntartja a szűk közegének. Arra már nincs érzelmi kapacitásunk, hogy vadidegen embereknek segítsünk. Miért is tennénk? Ez az okfejtés érthető. A mai felnőttek, akiknek az érzelmi és testi kapacitásainak nagy része az önfenntartásra megy el, és állandó harcot kell vívniuk a fennmaradásáért, sokszor már nem éreznek különösebb empátiát mások iránt, mert úgy gondolják, irántuk se érzékenyebbek mások. A „szabad felhasználású" empátia tehát a nyilvános helyzetekben elhalványul, marad a közöny, amivel elfordítjuk a fejünket. Nem kell csodálkoznunk akkor, ha ebben az individualista kultúrában ez a minta egyben csoportnorma is, és a felelősség megoszlása (majd más valaki biztos segíteni fog) következtében az agresszió enyhébb vagy erőteljesebb formában, de jelen van.

A személyiségfejlődés során mindenki átéli a különböző életkorokhoz kapcsolható agresszív érzéseket. Az optimális pszichés fejlődés, a nevelés és a felnőttek mintaadása, az anyai kapcsolat és a biztonságos elfogadás következtében, azt jelenti, hogy az agresszió nyílt levezetése gátolttá válik, tehát munkahelyi büfénkben nem két pofon kíséretében szerezzük meg a másik kávéját, hanem szépen kivárva a sorunkat, illedelmesen rendelünk magunknak.

Attól azonban, hogy a felszínen szép nyugalom és az indulatok jó kezelése látszik, rejtetten megmaradhat az agresszió vágya, ami legtöbbször tudattalan működésű, de állandóan jelen lévő tényező. A pszichoanalitikus elméletek alapján tudjuk, hogy minden személyiségben jelen vannak agresszív érzések, egészen korai életkortól kezdve. Alapvető fontosságú a legkoraibb érzelmi kapcsolatok kialakításában, de megtaláljuk a felnőttkori önérvényesítési képességben is.

A személyiség agresszív tendenciáinak megengedett, legalizált formáit és mértékét a nevelés folyamán közvetítik a szülők a gyereknek, ezt később a különböző intézmények kiegészítik, gyakorlatilag közvetítve a társadalmilag elfogadott konvenciókat. Ebben a folyamatban alakul a gyermek személyisége úgy, hogy a megengedhetetlen késztetések elfojtásra kerülnek, amelyek felett a szabályoknak, normáknak megfelelő módon a szülői minta alapján kifejlődött felettes én uralkodik. Annak, hogy a szülői minta milyen és hogyan formálja a gyereket, a legelső

állomása a korai anya-gyerek kapcsolat. A kötődés olyan érzelmi jelenség, amely alapvető a kapcsolatok kialakulásának a szempontjából. Az anya viselkedése és érzelmeket visszatükröző funkciója a legfontosabb tényezője ennek a folyamatnak (Hartmann 1999: 115).

Az, hogy egy kamasz hogyan és milyen mintázatú érzelmekkel reagál majd, mennyire képes kezelni saját és mások agresszióját, ezektől a korai érzelmi folyamatoktól függ. Ezek az alapok. A későbbiekben az, hogy a külvilág (szélesebb értelemben, mint az anya személye, tehát a felnőttek és a reális/virtuális tér szereplői) milyen mértékű agressziót mutat, és azt hogyan közli, vagyis mennyire teszi legitimmé annak megjelenését, szintén befolyásoló tényezőnek számít.

Az empátia fontos emberi képesség, „beleérzés", „mely által képesek vagyunk egy másik személy helyzetének az átélésére, miközben tudatában maradunk saját identitásunknak" (Rycroft 1994: 78). A nevelés során empatikus szerepmodellek kellenek: velük való interakciók és tőlük kapott visszacsatolások. A fegyelem, amit hagyományosan a szülőktől tanulunk, megköveteli, hogy a gyerek elképzelje, mit érezhet egy áldozat. Ezek az „empátiaalapú" forgatókönyvek aktiválódnak a hétköznapokban is, és befolyásolják a viselkedést. Amikor egy anya azt mondja a kisgyerekének, hogy ne rángassa a macit, vagy ne nyomja ki a baba szemét, akkor indoklásul elhangzik: „ne csináld, ez nagyon fáj neki". Ezen a módon kezdi tanítani az empátiát, vagyis a beleérzés képességét, arra ösztönözve gyermekét, hogy mielőtt fájdalmat okozna, gondolja végig annak következményeit.

Ma azonban a képmegosztó alkalmazások ontják magukból az agresszív tartalmakat, ami könnyen átcsúszik abba a tartományba, ahol „tanult érzéketlenséget" kiváltva, csak arról szól: „kemény vagyok, mert meg tudom nézni!" Azonban ez a felszín. Az érzelmi apparátus minden egyes alkalommal feldolgozza a látottakat, és válaszreakciót is indít.

A virtuális tér agresszív tartalmai – gondoljunk a TikTok-kihívásokra, amelyekben nem számít a fájdalom, a veszélyek – ma már mindennaposak. Pszichológiai tény, hogy az ilyen tartalmak megtekintése, utánzása olyan narcisztikus élmény, amely csábításának való ellenálláshoz komoly énerőre lenne szükség. Valójában agresszív verseny ez a javából, ahol nemcsak arról szól minden, hogy „minden" megmutatható, hanem arról is, hogy ki kell tűnni a többiek közül, olykor bármi áron. Az érzelmi háttér, amelyen ezen tartalmak megszületnek, és aztán követőkre találnak, valójában olyan tudattalan csoporthatás, amely észrevehetetlenül sodorja bele a kisebb ellenállással rendelkezőket.

Egyre több bizonyíték utal arra (Choukas-Bradley et al. 2022; Wick–Keel 2020), hogy az Instagramnak hatása van arra, hogy az emberek miként érzékelik magukat és a körülöttük lévő világot. Befolyásolja a felhasználók életérzését, a saját világukhoz való attitűdjüket, az önmagukról kialakított képet, egyszóval az önértékelést, az önbizalmat és a környezettel szembeni attitűdöt. A megosztások révén tulajdonképpen ez a folytonos híradás olyan közléssé válik, aminek már csak az a

lényege, milyennek látszunk. És miután az online személyiség működés valójában erodálja az én-funkciókat, a „belső előállítású" nyugalom és biztonság hiányában a folytonos én-prezentálás biztosítja majd a jóérzéseket.

Tulajdonképpen megtehetjük, hogy tágabb kontextusban a fogyasztói társadalmat olyan rideg szülőként értelmezzük, aki elfordult a „gyerektől" (személyiség), aki ilyenformán magányra és individuális létre van kárhoztatva, így lesz egyre nagyobb szüksége az online térbeli identitására, kapcsolataira és az ott átélhető érzelmekre. Ebben az individuális létben aztán „pótszerek" felé fordul, amelyeknek az öröm és a nyugalom állapotát kellene megteremteniük.

9. RIVALIZÁCIÓS DÜH ÉS A TÖNKRETÉTEL BOLDOGSÁGA

Régen a híres személyeket (színészeket, sportolókat, tudósokat, zenészeket, művészeket) tisztelet övezte. Nehezen lehetett megközelíteni, legfeljebb levelet lehetett írni neki, reménykedve abban, hogy kézhez kapja, és talán kinyitja a borítékot. Tekintélyszemélyek voltak ők, akik tehetségük vagy tudásuk okán kicsit felette álltak az átlagembernek. Senkinek nem volt baja ezzel a hierarchikus renddel, mert azt nagyjából mindenki tudta magáról, hogy képtelen lenne elvezényelni egy szimfonikus művet vagy eljátszani színpadon egy nagyjelenetet, esetleg írni egy verset, netán megnyerni egy olimpiát. Ez a távolság az átlagember és a híresség között nem mesterségesen megalkotott különbség volt, hanem ösztönös. Ha egy színházbejáróná összefutott a néző és a híres színész, akkor kellett néhány pillanat, amíg a civil összeszedte magát annyira, hogy meg merje szólítani a művészt. Az információs korban ez az érzelmi működés megváltozni látszik, legalábbis az online térben.

Ma a youtuberek alakítják a kultúrát. Ennek a kulturális *Zeitgeist*nek (korszellemnek) fő vonása, hogy a youtuberek meghallgatják a rajongóikat, együttműködnek velük, és olyan közösséget alkotnak, amely inkább barátságnak, semmint rajongói viszonynak tűnik. A tizenévesek nagy része inkább ezekkel az alkotókkal találkozik, semmint a hagyományos vagy professzionális hírességekkel. Ennek oka többet között az, hogy azt érzik, kedvenc alkotóik jobban megértik őket, mint a barátaik, és miközben alakítják a trendeket, annak ők is a részeseivé válnak.

Lélektani értelemben a tekintélyszemélyek által (is) kialakított régi tisztelet-hierarchia erodálása mögött van indulat. A hagyományok lebontása nem csak a média megváltozásában vagy a Zeitgeist formálódásában jelenik meg. Kell, hogy legyen mögötte valamilyen érzelmi hatás, ami milliók szemében emeli a hajdani átlag alattit is a magas fölé. Úgy véljük, ez a hatás nem más, mint a harag. Amikor azt írjuk, „infantilis harag", arra a lélektani pozícióra gondolunk, amely az online térben még erősebbé válik. Az azonnaliság, a hiperkonnektivitás lehetősége – még felnőtt életkorban is – csökkenti az érzelmi reakciók idejét, egyre gyorsabbá vál-

lényege, milyennek látszunk. És miután az online személyiség működés valójában erodálja az én-funkciókat, a „belső előállítású" nyugalom és biztonság hiányában a folytonos én-prezentálás biztosítja majd a jóérzéseket.

Tulajdonképpen megtehetjük, hogy tágabb kontextusban a fogyasztói társadalmat olyan rideg szülőként értelmezzük, aki elfordult a „gyerektől" (személyiség), aki ilyenformán magányra és individuális létre van kárhoztatva, így lesz egyre nagyobb szüksége az online térbeli identitására, kapcsolataira és az ott átélhető érzelmekre. Ebben az individuális létben aztán „pótszerek" felé fordul, amelyeknek az öröm és a nyugalom állapotát kellene megteremteniük.

9. RIVALIZÁCIÓS DÜH ÉS A TÖNKRETÉTEL BOLDOGSÁGA

Régen a híres személyeket (színészeket, sportolókat, tudósokat, zenészeket, művészeket) tisztelet övezte. Nehezen lehetett megközelíteni, legfeljebb levelet lehetett írni neki, reménykedve abban, hogy kézhez kapja, és talán kinyitja a borítékot. Tekintélyszemélyek voltak ők, akik tehetségük vagy tudásuk okán kicsit felette álltak az átlagembernek. Senkinek nem volt baja ezzel a hierarchikus renddel, mert azt nagyjából mindenki tudta magáról, hogy képtelen lenne elvezényelni egy szimfonikus művet vagy eljátszani színpadon egy nagyjelenetet, esetleg írni egy verset, netán megnyerni egy olimpiát. Ez a távolság az átlagember és a híresség között nem mesterségesen megalkotott különbség volt, hanem ösztönös. Ha egy színházbejáróná összefutott a néző és a híres színész, akkor kellett néhány pillanat, amíg a civil összeszedte magát annyira, hogy meg merje szólítani a művészt. Az információs korban ez az érzelmi működés megváltozni látszik, legalábbis az online térben.

Ma a youtuberek alakítják a kultúrát. Ennek a kulturális *Zeitgeist*nek (korszellemnek) fő vonása, hogy a youtuberek meghallgatják a rajongóikat, együttműködnek velük, és olyan közösséget alkotnak, amely inkább barátságnak, semmint rajongói viszonynak tűnik. A tizenévesek nagy része inkább ezekkel az alkotókkal találkozik, semmint a hagyományos vagy professzionális hírességekkel. Ennek oka többet között az, hogy azt érzik, kedvenc alkotóik jobban megértik őket, mint a barátaik, és miközben alakítják a trendeket, annak ők is a részeseivé válnak.

Lélektani értelemben a tekintélyszemélyek által (is) kialakított régi tisztelet-hierarchia erodálása mögött van indulat. A hagyományok lebontása nem csak a média megváltozásában vagy a Zeitgeist formálódásában jelenik meg. Kell, hogy legyen mögötte valamilyen érzelmi hatás, ami milliók szemében emeli a hajdani átlag alattit is a magas fölé. Úgy véljük, ez a hatás nem más, mint a harag. Amikor azt írjuk, „infantilis harag", arra a lélektani pozícióra gondolunk, amely az online térben még erősebbé válik. Az azonnaliság, a hiperkonnektivitás lehetősége – még felnőtt életkorban is – csökkenti az érzelmi reakciók idejét, egyre gyorsabbá vál-

nak bizonyos érzelmi szükségletkielégítési igények, amelyeket az online tér látszólag fogad és válaszol rájuk.

Az egyre gyorsabb készülékek könnyű elérése, az azonnali érzelmi szükségletkielégítés lehetősége, az érzelmi inkontinencia (azonnali megosztások) nem az érett személyiség működésének a jellemzőit erősítik, hanem éppen ellenkezőleg, olyan infantilis attitűdöt emelnek a mindennapi normalitás keretei közé, amelynek eredményeképpen látjuk a rugalmatlanságot, a haragot, az indulatokat, a felelősség áttolásának lehetőségét, de legfőképpen a megdolgozatlan és fésületlen impulzusok cseréjét.

A negyedik ipari forradalom termékei, a filterek és a valós idejű fordítóalkalmazások, az applikációk, melyeken zenét szerezhetünk, festhetünk, fotózhatunk – mind illúziót teremtenek, és azt ígérik: tehetséges művészek lehetünk. Miközben csak akkor lehetünk valóban azok, ha célokat alkottunk, erőfeszítéseket tettünk, ha kudarcot vallottunk, tanultunk a hibáinkból, és újrakezdtük. De a digitális fejlesztések most elhozták azokat a lehetőségeket, amelyek megtévesztőek, mert elhitetik, hogy mindenkiből lehet VALAKI. Ez pedig az átlagemberben is felélesztí a vágyat, hiszen a narcisztikus igények végtelenek, különösen, ha könnyen elérhetőnek látszanak.

10. ÖSSZEFOGLALÁS

Korábban sosem kellett együtt élnünk ilyen erős és önálló nem emberi rendszerekkel. A legnagyobb kockázat, hogy nem leszünk hajlandók vagy éppen készek arra, hogy kritikusan gondolkozzunk azokról a változásokról, melyeket magától értetődőnek kezdünk tekinteni. Egyre nehezebb lesz tehát, hogy különbséget tegyünk a valóság és a virtualitás között.

Harari megállapítása szerint: „Az egyén zavarbaejtően keveset tud a világról, és a történelem előrehaladásával tudása egyre csak csökken. Egy kőkorszaki vadászó-gyűjtögető emberhez képest mi szinte valamennyi szükségletünk kielégítéséhez mások tapasztalatait vesszük igénybe. Noha egyénileg nagyon keveset tudunk, azt hisszük, hogy nagyon is sokat, mivel a mások elméjében lévő tudást is úgy kezeljük, mintha a sajátunk volna. Evolúciós szempontból a Homo Sapiensnek remekül bevált az, hogy mások tudásában bízott. De [...] a tudásillúziónak is van árnyoldala. A világ egyre komplexebb lesz, és az emberek nem képesek felmérni, milyen keveset tudnak róla. Ennek eredményeképpen a hasonlóan gondolkodó barátok és önmegerősítő hírfolyamok visszhangkamrájába zárkóznak, így aztán elképzeléseik folyamatos alátámasztást kapnak, viszont csak ritkán kérdőjelezik meg őket" (Harari 2018: 194).

Számít tehát, hogy mennyire engedjük el azt a gyeplőt, amely az érzelmi működésünk kontrollját jelentené. Most kell megtanulnunk, hogyan kell óvatosan

és okosan élni ennek az új világnak az eszközeivel, hogy ne emberi kapcsolataink érzelmi minőségén csapódjon le egy eddig ismeretlen hatás.

Míg a biológiai érés továbbra is múlt századi, és a pszichés, érzelmi funkciók biológiai alapjai továbbra is a múlt századi léptékekben fejlődnek ki, addig az információtechnológia fejlődése felgyorsít bizonyos kognitív funkciókat és érzelmi működéseket, amelyekhez azonban – főleg a digitális generációknál – a biológiai korhoz rendelhető feldolgozó kapacitás nem áll még rendelkezésre. Harari állítása szerint: „A veszély abban rejlik, hogy ha túl sokat fektetünk az MI, és túl keveset az emberi tudatosság fejlesztésébe, előfordulhat, hogy a számítógépek rendkívül kifinomult mesterséges intelligenciája csupán az ember természetes ostobaságának hatalomra jutását szolgálja majd. Míg a tudományos-fantasztikus thrillerek drámai, lánggal és füsttel teli apokalipszist ábrázolnak, a valóságban inkább egy banális, kattintgatós apokalipszis várhat ránk. Ennek elkerülése érdekében bölcs lenne minden dollár és perc mellé, amit a mesterséges intelligencia fejlesztésébe fektetünk, befektetni egyet-egyet az emberi tudatosság fejlesztésébe is" (Harari 2018: 71).

Valójában tehát az a paradox helyzet áll elő, hogy a mesterséges intelligencia korában, önmagunk védelmében olyan öröktől jelen lévő képességeinkre kell majd támaszkodnunk, mint a józan ész, a tudatosság, a megérzések. Ahogy láttuk, agyunk képes arra, hogy megkülönböztessen valódi és nem valódi jelenségeket. Erre érdemes koncentrálnunk ahelyett, hogy elhitetjük magunkkal: urai vagyunk a helyzetnek.

El fog jönni az az idő (ha még nem jött el), amikor egy szakember sem tudja megmondani, hogy valós vagy nem valós videót lát-e. Ez azt jelenti, hogy eljuthatunk egy másik pontig is: ez a valóságosba vetett hit megingása, amikor alapvető hozzáállásunkká válhat, hogy már nem hiszünk annak, amit látunk, vagy már nem is tudjuk, hogy minek hihetünk egyáltalán.

SZAKIRODALOM

Aronson, Ellis – Tavris, Carol 2009: *Történtek hibák (de nem én tehetek róluk).* Budapest: Ab Ovo.

Choukas-Bradley, Sophia – Roberts, Savannah R. – Maheux, Anne J. – Nesi, Jacqueline 2022: The Perfect Storm: A Developmental–Sociocultural Framework for the Role of Social Media in Adolescent Girls' Body Image Concerns and Mental Health. *Clin Child Fam Psychol Rev*, 25/4: 681–701. doi: 10.1007/s10567-022-00404-5

Course-Choi, Jenna – Hammond, Linda 2021: Social media use and adolescent well-being: A narrative review of longitudinal studies. *Cyberpsychology, Behavior, and Social Networking*, 24/4: 223–236. doi: 10.1089/cyber.2020.0020.

Miklósi Ádám 2010: Kitekintés a jövőbe. In: Csányi Vilmos – Miklósi Ádám (szerk.): *Fékevesztett evolúció.* Budapest: Typotex. 161–169.

Moshel, M. L. – Robinson, A. K. – Carlson, T. A. – Grootswagers, T. 2022: Are you for real? Decoding realistic AI-generated faces from neural activity. Vision Research 199/108079. https://doi.org/10.1016/j.visres.2022.108079

Eco, Umberto 2008: *Az új középkor.* Budapest: Európa Kiadó.

Fodor Péter – L. Varga Péter 2018: Marshall McLuhan. In: Kricsfalusi Beatrix – Kulcsár Szabó Ernő – Molnár Gábor Tamás – Tamás Ábel (szerk.): *Média- és kultúratudomány.* Budapest: Ráció Kiadó. 447–454.

Guld Ádám 2022: Bevezető tanulmány. In: Szadai Károly (szerk.): *VV10 – Egy valóságshow valósága.* Budapest: Médiatudományi Intézet. 9–20.

Harari, Yuval Noah 2018: *21 lecke a 21. századra.* Animus.

Hartmann, Ellen 1999: Lichtenberg: szelfpszichológus vagy motiváció-teoretikus? In: Karterud, Sigmund – Monsen, Jon T. (szerk.): *Szelfpszichológia – a Kohut utáni fejlődés.* Budapest: Animula Kiadó. 110–122.

Keen, Andrew 2007: *The Cult of the Amateur, How today's internet is killing our culture and assaulting our economy.* New York: Crown Business.

Konok Veronika – Peres Krisztina – Ferdinandy Bence – Jurányi Zsolt – Bunford Nóra – Ujfalussy Dorottya Júlia – Réti Zsófia – Kampis György – Miklósi Ádám 2020: Hogyan hat a mobileszköz-használat az óvodások figyelmére és társas-kognitív készségeire? *Gyermeknevelés*, 8/2: 13–31.

Köbis, Nils – Starke, Christopher – Soraperra, Ivan 2021: The Psychology of Deepfakes. People can't reliably detect deepfakes but are overconfident in their abilities. *Psychology Today*, december 2. https://www.psychologytoday.com/intl/blog/decisions-in-context/202112/the-psychology-deepfakes

Köbis, N. C. – Doležalová, B. – Soraperra, I. 2021: Fooled twice: People cannot detect deepfakes but think they can. iScience 24/11. https://doi.org/10.1016/j.isci.2021.103364

McLuhan, Marshall – Fiore, Quentin 2012: *Médiamasszázs.* Budapest: Typotex.

Purnell, Carolyn 2020: Do We All Still Agree that "Seeing Is Believing"? Deepfakes destabilize our collective notions of truth. *Psychology Today*, június 23. https://www.psychologytoday.com/us/blog/making-sense/202006/do-we-all-still-agree-seeing-is-believing

Rycroft, Charles 1994: *A pszichoanalízis szótára.* Budapest: Párbeszéd Könyvek.

Somlai Péter 1997: *A szocializáció elmélete.* Doktori disszertáció, kézirat. n. a. Online: https:/www.fszek.hu/szociologia/somlai_peter_a szocializáció_elmelete.pdf

Tari Annamária 2013: *Ki a fontos? Én vagy Én?* Budapest: Tericum Kiadó.

Timeo, Susanna – Riva, Paolo – Paladino, Maria Paola 2020: Being liked or not being liked: A study on social-media exclusion in a preadolescent population. *Journal of Adolescence*, 80/1: 173–181. https://onlinelibrary.wiley.com/doi/10.1016/j.adolescence.2020.02.010

Tóth Péter István 2011: *A médiaerőszak-kommunikáció hatásának újraértelmezése az evolúciós viselkedéstudományok perspektívájából.* Doktori disszertáció, kézirat. Pécsi Tudományegyetem Bölcsészettudományi Kar Nyelvtudományi Doktori Iskola Kommunikáció Program.

Twenge, Jean M. 2018: *iGeneráció.* Budapest: Édesvíz Kiadó.

Vogel, Erin A. – Rose, Jason P. – Roberts, Lindsay R. – Eckles, Katheryn 2014: Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture*, 3/4: 206–222. https://doi.org/10.1037/ppm0000047

Wallace, Patricia 2006: *Az internet pszichológiája.* Budapest: Osiris.
Wick, Madeline R. – Keel, Pamela K. 2020: Posting edited photos of the self: Increasing eating disorder risk or harmless behavior? *International Journal of Eating Disorders*, 53/6: 864–872.

FORRÁSOK

Influencer Kisokos 2020: Hódítanak a virtuális influencerek – de kik is ők valójában? *Star Network*, május 22. https://influencerkisokos.starnetwork.hu/2020/05/22/hoditanak-a-virtualis-influencerek-de-kik-ok/
W1 = Deankadani 2021: 1 millió dollárt keres az Insta képeivel, pedig még csak nem is létezik. *Leet,* szeptember 15. https://leet.hu/2021/09/15/1-millio-dollart-keres-az-insta-kepeivel-pedig-meg-csak-nem-is-letezik/
Varga Csenge Virág 2022: A hangalapú okoseszközök akadályozhatják a gyermekek szociális és kognitív fejlődését. *Index,* szeptember 29. https://index.hu/techtud/2022/09/29/hangalapu-okoseszkozok-apple-amazon-google-home-siri-gyermekek-fejlodese-empatia-kognitiv-fejlodes-pszcihologia/ [sic]

KÁRPÁTI ANDREA

A deepfake hatása az oktatásra

A deepfake jelen van az iskolában, de a tananyag hitelességét még nem veszélyezteti. A fejezet a tudományos és művészeti területen megjelenő, az oktatás világára potenciálisan veszélyes hamis hírek néhány típusának bemutatásával kezdődik. „Mit tehet a tanár?" – ezzel a kérdéssel folytatódik a fejezet, amely az oktatás területén végzett, a deepfake elterjedtségét és hatását vizsgáló kutatások eredményei és külföldi jógyakorlatok bemutatásával érzékelteti, milyen nevelési elvek, szabályozó intézkedések segíthetnek megakadályozni az álhírek beépülését a fiatalok gondolkodásába. Jó példákat hozunk a bevallott, hiteles adatokon alapuló deepfake kreatív oktatási használatára. A fejezet végén megkérdezzük: „Mit tehet a szülő?" Ahhoz, hogy hatásosan cselekedhessenek, a felnőttek digitális kompetenciájának fejlesztésére, az etikus internethasználat elveinek elfogadására, a képmásokhoz fűződő személyiségi jogok megismerésére van szükség. Mindezeket a mai iskola megtanítja – a kérdés az, mennyire tanítható ugyanerre a társadalom.

Kulcsszavak: autentikus digitális tartalom, digitális tartalomhitelesítés, digitális írástudás, médiakompetencia, médiaetika

1. A DEEPFAKE MÁR AZ ISKOLÁBAN VAN?

Az iskolai tanterv és a hozzá kapcsolódó tananyag világszerte számos ellenőrzésen átesik, míg kötelező vagy ajánlott tudnivalóként a diákok elé kerül. A törzsanyagnak nevezett, Magyarországon központilag előírt, kötelező tudásanyag tartalma nehezen változik, és a kiegészítő és segédanyagokkal együtt számos lektori körön, kipróbálási fázison megy át. Úgy tűnik, a deepfake jelenleg még nem férkőzött be a tananyagba. *A tanulás tartalmát tehát még nem, az iskolai közösség morálját, a tanárok és diákok személyiségi jogait azonban már megjelenése óta veszélyezteti.* Egy jelentős, az állami iskolákat is ellenőrző kanadai digitális biztonsági cég, az eSafe munkatársa, Jordan Foster (2021) szerint a deepfake elsősorban a személyközi kommunikációban van jelen: az elektronikus zaklatás *(cyberbullying)* számára kínál hatásos és nehezen cáfolható álhírgyártási lehetőséget. Kanadában ezt a veszélyt olyan jelentősnek tartják, hogy külön kormánymegbízott foglalkozik az elektronikus hamisítási ügyekkel, és országos felvilágosító kampányt indítottak

ellene. Foster szerint a deepfake ezekben a formákban fedezhető fel az oktatási intézményekben:

- elektronikus zaklatás *(cyberbullying);*
- pszichológiai befolyásolás *(social engineering):* online üzenetváltások, amelyek veszélyes, káros, illegális cselekedetekre vagy érzékeny adatok kiszolgáltatására veszik rá a fiatalokat;
- egyéb visszaélés a képmásokkal *(image-based abuse).*

Gyakran esik szó a Z generáció (az 1990-es évek közepétől a 2010-es évek elejéig születettek) digitális kompetenciájáról, mivel ez az első nemzedék, amely tagjai már kisgyermekkorukban találkozhattak a digitális eszközök széles körével. A Magyarországon kötelező, külön tantárgyként oktatott digitális kultúra (korábban: informatika) tantárgy mellett a mozgóképkultúra és médiaismeret, illetve a rajz és vizuális kultúra tantárgyak programjában szerepel *a digitális kompetencia sajátos képességcsoportja: a digitális kreativitással, az álló és mozgóképek szöveggel és hanggal kísért, szabad, esztétikai igényű előállítása.* A digitális képalkotó technika a 21. századi gyermekrajz- és kamaszalkotás vezető műfaja, a fiatalok izgalmas, egyre gazdagabb tartalmú és egyre igényesebben megformált új képi nyelve (Kárpáti–Nagy 2019). Hogy mire használják a diákok a vizuális kommunikáció új lehetőségeit, elsősorban tanáraiktól, de szüleiktől is függ. Az etikus képalkotás és képhasználat elsajátítása éppolyan fontos feladatunk, mint a bevezetés a digitális alkotó eszközök használatába.

Az iskolában már ott van, az oktatást még nem érte el a deepfake, de gyorsan közeledik felé. Akkor fog hatni rá, ha több lesz gyorsan leleplezett blöffnél, és sikerül *tudományos narratívát építeni a hamisítvány köré.* Egy, jóval a deepfake előtti példával illusztrálható, milyen a hatásos, a hamisított alkotást a művészettörténetbe ágyazó tudományos narratíva. Henrikus van Meegeren a 20. század első évtizedeiben kezdte festői pályafutását Hollandiában, és hamar a 17. századi németalföldi mesterek stílusának hatása alá került (Lopez 2009). A képi nyelvet alapvetően megújító Picasso, Kandinszkij, Chagall vagy hazájában, Hollandiában a tiszta absztrakció stílusában alkotó de Stijl művészcsoport kortársaként, nagy sikerre nem számíthatott, műveit idejétmúltnak találták. A kritikák hatására alkotó módszert nem, csak pályát változtatott, és mint Jan Vermeer van Delft (1632–1675) képeinek legjobb hamisítója, óriási sikereket ért el. Egyik alkotását, az *Emmausi vacsorát*, amikor „csodás módon előkerült", a kritikusok a hamisító leleplezéséig a mester legjobb művének tartották.

Van Meegeren hamisítói sikerének titka az volt, hogy nemcsak a létező Vermeer-művekhez hasonló témákat választott, hanem *kitalált egy új alkotói korszakot, s ezzel egy új tudományos narratívát* is a németalföldi festészet egyik legjelentősebb mesterének: katolikus tematikájú képeket festett a nagy művész stílusában. A filmen is megörökített, Észak Mona Lisájának nevezett „Lány gyöngy fülbevalóval"

1. ábra. Van Meegeren a műhelyében, 1945 (fotó: Koos Raucamp; forrás: W2)

alkotója a felesége kedvéért lett reformátusból katolikus, de tiszteletben tartva a protestáns delfti városvezetés (legfőbb megrendelői) meggyőződését, képein csak nagyon ritkán és áttételesen ábrázolt katolikus hittételekre utaló jeleneteket. Van Meegeren hamisítványai tehát azért voltak különösen hatásosak, mert egy, Vermeer életrajzával megmagyarázható, de a festői életműből hiányzó alkotói korszakot épített fel belőlük (De Guise 2022). Hasonló módon, az oktatás tartalmára azok a deepfake-művek veszélyesek, amelyek új történelmi adatként vagy tudományos eredményként eddigi ismereteink közé könnyen beépíthetők, sőt új belátásokhoz vezetnek.

A pedagógus nem tehet többet, mint a műtárgybecsüs: alaposan megvizsgálja a szeme elé kerülő információt, annak eredetét (a műtárgypiac szóhasználatával: *provenanciáját*). Pontosan ezt a kifejezést használja céljainak meghatározására a Tartalom Eredetiségét Vizsgáló Hatóság (*Content Authenticity Initiative*, W1), amely a Twitter, az Adobe, a New York Times, a BBC és más jelentős médiacégek összefogásával alakult meg. Célja a publikált digitális tartalom eredetének (provenanciájának) vizsgálata, az információ eredetének felderítésével.

A szöveges és képi adat és alkotás hitelessége ellenőrzési módjainak a digitális kompetencia fejlesztésével foglalkozó tantárgyakon kívül immár *valamennyi tan-*

tárgy anyagában szerepet kell kapniuk. A deepfake legnagyobb veszélye a tanításban az, hogy beépülve a valódi tények, adatok és gondolatok közé, hamis irányba terelheti a fiatalok gondolkodását. Bizarr módon a deepfake elnevezés magában foglalja a tanulásra való hivatkozást, hiszen a kifejezés első szava a mélytanulásra *(deep learning)* utal. A mesterséges intelligencia segítségével alapos szemlélés után is hitelesnek tűnő álló- vagy mozgókép készíthető például egy történelmi jelentőségű találkozóról vagy tudományos felfedezésről, amely az iskolai tananyagban szerepel, és amelyhez a gyanútlan tanár felhasználhatja a hamis adatokat tartalmazó „dokumentumot". A *publikálás kontextusa* egyelőre még segíti a hitelesség megállapítását, de egyre gyakrabban lepleznek le az egyetemek vagy kutatóintézetek, politikai szervezetek vagy a műtárgypiac szereplői digitális megjelenését utánzó, bár más neveket használó, patinásnak tűnő portálok, amelyek alkalmasak arra, hogy alátámasszák az általuk kommunikált információk hitelét (Guld 2021).

„A legalapvetőbb változás az, hogy az igaz és hamis megkülönböztetésének felelőssége immár a mi egyéni feladatunk. Az az elképzelés, hogy társadalmi kapuőrök fogják megmondani nekünk, hogy egy dolog igaznak bizonyult, immár romokban hever" (Kovach–Rosenstiel 2010: 7).[1]

Aczél Petra az álhírről mint társadalmi jelenségről szóló tanulmánya (2017) bevezetőjében szerepel ez az idézet, amelyben *a „társadalmi kapuőrök" kifejezést az oktatásban a pedagógussal azonosíthatjuk.* Aczél egy 17. századi kommunikációs kampányról számol be, amelynek célja az volt, hogy a XIV. Lajos által megalapított Tudományos Akadémia munkáját megismertesse és elfogadtassa a francia közvéleménnyel. A kampányban a nagyobb hatás kedvéért meghamisított képeket használnak fel.

A mesterséges intelligencia *(artificial intelligence, AI)* további veszélyt jelent az oktatásban: a tanulók teljesítményének megítélése is veszélybe kerülhet. Immár nemcsak a tananyagban szereplő csak a tényeket és műveket, hanem a nyitott, kreatív megoldásokat igénylő házi feladatokat is lehet hamisítani. A digitális eszközökkel végezhető csalás a zárthelyi, írásbeli, teszt jellegű feladatoknál a mobiltelefonokkal egyidős probléma, hiszen a megoldások ezzel az eszközzel nemcsak kikereshetők a webes adatbázisokból, de bármiféle tudásszerző munka nélkül, egy szakértő ügyesen leplezett felhívásával meg is hallgathatók. A mesterséges intelligenciára épülő nyelvi modellek, például az OpenAI API-nak (nyitott mesterséges intelligenciára épülő alkalmazásprogramozási felületnek) nevezett eszközbe már a vadonatúj GPT-3 nevű generatív nyelvi modellt is beépítették. Ez a szoftver képes arra, hogy megközelítse az emberek szövegértési és fogalmazási képességeit, vagy-

[1] „The most fundamental change is that more of the responsibility for knowing what is true and what is not now rests with each of us as individuals. The notion that a network of social gatekeepers will tell us that things have been established or proven is breaking down" (Kovach–Rosenstiel 2010: 7).

is olyan szövegeket hozzon létre tetszőleges témáról, amelyeket egy tanár nem tud megkülönböztetni egy diákja fogalmazványától.

Bodnár Zsolt (2022) számol be egy házi dolgozatokat gyártó csalássorozatról, amelyet büszke elkövetője, egy norvég diák a Redditen tett közzé. A 16 éves norvég esszégyáros szerint az AI az érvelő prózában is kiválóan teljesít, a tanárok jó jegyekkel ismeri el az esszéket. A tanároknak szóló subredditen saját kibeszélőt kapott ez a téma, és a leginkább vigasztaló érv az volt, hogy az AI-technológia még nem tart ott, hogy egy informatikában alapszinten képzett felhasználó az emberi gondolkodást és stílust követő szövegeket hozzon létre. Ha a tanár összevetette volna a hamisított szöveget egy-két korábbi, eredeti fogalmazvánnyal, könnyen felfedhette volna a csalást. A vitatkozók közül azonban senki nem kételkedett abban, hogy az AI fejlődni fog, és a tanároknak rövid időn belül ezzel a csalási módszerrel is számolniuk kell. Emellett szól a DeepMind szoftver remek teljesítménye az USA országos, középiskolai szövegértési tesztjének megoldásában (ennek nyelvészeti és informatikai hátteréről vö. Rae et al. 2021). Az egyetlen hatásos ellenőrzési mód a jövőben a hagyományos feleltetés lesz: a gyanús esszébe foglalt tudásanyag kikérdezése az állítólagos szerzőtől, az osztályteremben.

Az oktatás számára a deepfake-hírek rohamos terjedése jelenti az igazi kihívást. Ha az álhír korábban megbízhatónak elismert portálokon is megjelenik, tehát olyan orgánumokig elér, amelyet a tanítványok és szülők (és sok esetben a tanár is) hitelesnek fogadnak el, nagyon nehéz a hamar kialakuló és az álhírrel alátámasztott tény megcáfolása (vö. az álhírről Veszelszki 2021 elemzését). A Z generáció számos olyan digitális kép- és szöveghasználati módot tett mindennapi kommunikációs gyakorlattá, amelyekben a hamis információ központi szerepet kap. A cyberbullying és a szelfikultúra egyaránt a testkép meghamisításáról szól, az influenszerkommunikációban vagy a realityműsorokban számos, hitelesnek beállított, szándékos vagy a tetszés igényéből fakadó, igaznak hitt hibás állítás kap óriási nyilvánosságot.

A képátalakítás célja persze nem mindig a félrevezetés. A nyíltan szórakoztató célú képfeldolgozásokat a tanárok egy része a *digitális kreativitás fejlesztésére* az oktatásban is alkalmazza. Ilyenek például a híres festmények és fotók alapján készült sorozatok, amelyek a modellt sokféle arckifejezéssel és nézetből ábrázolják vagy Van Gogh tájképeihez készítenek meghökkentő animációt, a képmezőből a néző felé felszálló varjakkal, hunyorgó csillagokkal. Egyelőre még többen vannak azok, akik a műalkotások megcsúfolását látják az ilyen képi kísérletekben.

Ezek az alkotások szándékuk szerint nem a deepfake körébe sorolhatók, hiszen becsapni nem akarnak senkit, és műfajuk a *parafrázis:* egy létező alkotás átfogalmazása. Mégis érdemes az oktatással kapcsolatos tudásanyag-manipulációk között szólni róluk, hiszen alkalmasak arra, hogy a tanulók ezek alapján ítéljék meg az eredeti művet, félreértelmezve annak hangulatát és üzenetét. Aki az olajfestmény helyett vibráló színfoltok, cikázó fénysávok és vészesen közeledő, fekete madarak látványával találkozik először, könnyen szegényesnek érzi az eredeti kép

2. ábra. Állókép alapján generált deepfake művek (forrás: Barber 2019)

3. ábra. Animáció Vincent Van Gogh „Csillagos égbolt" című műve alapján (a kép adatai: olajkép vásznon, 1889, New York, Modern Művészeti Múzeum; forrás: W3)

finom, elmélyülést kívánó képi eszközeit. Aki egy festmény központi figuráját egy gifen látja meghökkenni, fintorogni vagy döbbenten meredni a semmibe, ezeket az emlékké vált képeket óhatatlanul rávetíti majd az eredeti műalkotáson megörökített arcra. Nem biztos, hogy az összevetésben a remekmű lesz a győztes, és az elmaradt élményt nehezen írja majd felül a művészről szerzett tudás.

2. MIT TEHET A TANÁR?

A *médiatudatossággal* immár az oktatás minden területén foglalkozni kell. Az új magyar informatikai tanterv bőven tartalmaz a médiafogyasztás veszélyeivel foglalkozó ismereteket, de a digitális írástudás *(digital literacy)* fejlesztése nem lehet egyetlen tantárgy feladata. Maga a kompetencia is sokat változott, gazdagodott az elmúlt két évtizedben. Része lett az információs írástudás *(information literacy)*, vagyis a sokféle forrásra támaszkodó információgyűjtés, de a reprodukciós írástudás *(reproduction literacy)* is, a hiteles helyekről beszerzett információk továbbadásának képessége. Külön képességként határozzuk meg a kontextualizálást *(branching literacy)*, amely az információ hátterének feltárását, magas szintű értelmezést jelent, illetve a vizuális képességrendszert *(photo-visual literacy)*, amely a képi információk szakszerű értelmezését jelöli. Ezek a kommunikációban alapvető képességek a 21. században a digitális írástudás támogatásával működnek (Blankenship 2021).

Eileen Dombrovski egyike volt azoknak a pedagógusoknak, akik a deepfake megjelenésekor rögtön felhívták a figyelmet arra, hogy a tények és adatok manipulálása jóval több, mint ártatlan vagy sértő tréfa.

> Elképzelhető-e, hogy a deepfake-nek nevezett mesterséges intelligencia tényleg társadalmi nyugtalanságot, politikai zűrzavart, nemzetközi feszültséget okoz, és még háborúhoz is vezethet? Lehetséges, hogy korábbi módszereink, amelyekkel felismertük, mi igaz, és mi hamis, végérvényesen használhatatlanná váltak? Vajon mi, a tanárok, lefelé robogunk a lejtőn, sebesen értékelve az elénk kerülő tényeket és adatokat, és igyekszünk időben irányt váltani, miközben vezetni tanítjuk a diákjainkat?
>
> A technikai fejlődés elképesztő iramot diktál, kritikai gondolkodásunkat pedig újabb és újabb kihívások érik. A deepfake nem egyszerűen a képek és videók jobbítására tett, jelentéktelen kísérlet. Ezek az alkotások a gépi tanulás fejlődésének példái. A mesterséges intelligencia tanul, amikor egy algoritmus segítségével képessé teszi a felhasználót arra, hogy olyan képelemeket építsen be egy videóba, amelyek az eredetiben nem voltak jelen. [...] Az önök diákjai gyorsan felismerik majd, mire lehet használni ezt a tudást, amint észlelik, hogyan fordul ellenük vagy mások ellen (Dombrovski 2018: o. n.).[2]

[2] „Could the development in artificial intelligence dubbed 'deepfakes' really »trigger social unrest, political controversy, international tensions« and »even lead to war«? Have our previous methods of telling fact from fiction been irremediably undermined? As teachers, we're careening down new paths in evaluation of knowledge claims, trying to learn to steer in time to teach our students to drive! Technology just got even more amazing, and our everyday critical thinking just got even more challenging. »Deepfakes« are not merely a mini-advance in digital adjustment of images and videos. Instead, they are developments in machine learning,

Tanítványaink technikai kihívásnak tekintik, mi elsősorban a tartalomhamisítás tudásrelativizáló hatását és ennek csökkentési lehetőségeit kutatjuk. A legnagyobb kihívás az oktatás számára a tudományosan igazolt tényeket cáfoló álinformációk azonosítása. A számos, hitelesnek tűnő, vonzó forrásból megerősített álhír, hamis tény ellen az oktatás csak úgy védekezhet, ha hasonlóan vonzó formában, sokféle információs csatornát felvonultatva mutatja be a cáfolatot. Dombrovski (2018) úgy véli, a pedagógusnak továbbra is erősnek és határozottnak kell mutatkoznia, nem terjesztheti a tények relativizálódásának az igazság utáni korra jellemző gondolatait. Azt tanácsolja az oktatóknak, hogy továbbra is a tények tiszteletére neveljenek. *A tanárnak tudásátadó szakemberként kell viselkednie*, aki kész rá, hogy megküzdjön a hamisítókkal, elsajátítsa a deepfake azonosításának emberi tapasztalatokon alapuló módszereit, de a szoftveralapú felismerési lehetőségeket is.

A projektmódszer, amelyben – a tanár mentorálásával – a fiatalok maguk erednek egyes jelenségek nyomába, és lehetőségük van *felfedezni, a társakkal megvitatni, majd a tanár és más szakértők elé tárni egy, számukra is fontos téma sokféle olvasatát*, éppen ilyen, közösségi tudásszerzésen alapuló, kritikus ismeretszerzést modellez. A diákok az információkereső munka kezdetén megismerhetik azokat a kutatási eredményeket, amelyek alkalmasak az álhírek elkülönítésére a valódi tudományos tartalmaktól. Veszelszki Ágnes az álhírek olyan nyelvi és vizuális jellemzőit tárta fel, amelyeket egy diák is jól azonosíthat kutatómunkája közben: a szenzációhajhász címet, a bulvármédiára jellemző, túlzásoktól hemzsegő szöveget, a szakkifejezések és a központozás helytelen használatát, a feltűnő képeket, a valódi oldalakat utánzó URL-címeket stb. (Veszelszki 2017). Hasonló útmutató készült az audiovizuális megtévesztő tartalmak felismerésével kapcsolatban (Veszelszki et al. 2022). Ha a tanár felhívja a figyelmet az információk megerősítésének szükségességére, és a diák más forrásokat is keres egy-egy meglepő, a korábbi ismeretekkel ellentétes hír hitelesítésére, a deepfake-tartalmak jelentős része leleplezhető. Egy hamis hír köré lehet történetet szőni, de az igazoltan hiteles hírportálokon ez remélhetőleg nem fog megjelenni, hiszen a szerkesztők alaposan ellenőrzik a megosztandó tartalmakat.

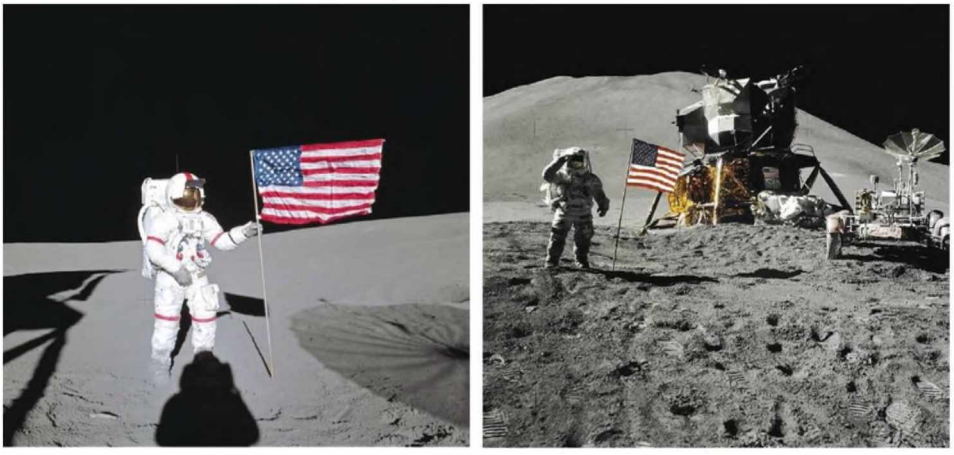
A hamis híreknél lényegesen nehezebben felismerhetők a deepfake-technológiával készülő videók, amelyekben egy filmrészlet szereplőit egy másik személy arcával és hangjával helyettesítik. A deepfake-videók vizuális megjelenése az elmúlt években gyorsan javult, s ma már egy átlagos felhasználó alig vagy egyáltalán nem tudja megállapítani, hogy valóságos tartalmakat vagy digitális ha-

---

as artificial intelligence learns and applies the algorithms to enable users to replace elements of a video with other ones not part of the original. (…) Your students will be quick, I'm sure, to imagine possible uses of this technology if directed against them or against others" (Dombrovski 2018: o. n.).

misítványokat lát-e. Különösen veszélyesek lehetnek a politikusok lejáratására készült videók, amelyek képesek megváltoztatni egy személyről hiteles megnyilvánulásai alapján kialakult képet. Ráadásul egy közszereplő bármikor megtagadhatja és hamisítványnak nevezheti korábbi megnyilvánulásait, ha már nem tartja vállalhatónak őket. *A deepfake káros hatásai közé tartozik a bizalomvesztés a médiában közölt információban* (Guld Ádámot idézi Bencze 2022). Az oktatás alapja pedig a bizalom: amit a tanár bemutat, az hasznos, értékes és hiteles tudás. A társadalomtudományok oktatásakor gyakran használnak dokumentumfilm-részleteket, ezért fontos, hogy a tanár tudatában legyen a deepfake-ben rejlő hamisítási lehetőségeknek. Ha bemutatás előtt ellenőrzi, hogy más, egyértelműen hiteles forrásból is igazolható-e, amit a videó bemutat, csak akkor tárhatja a filmet a diákjai elé.

A hírhamisítók leleplezései a valódi kutatási eredményekben való kételkedéshez is vezettek. A tudomány eredményeit százötven éve népszerűsítő *Popular Science* oldal a Holdra szállásban kételkedők meggyőzésére hétlépéses ellenőrzési módszert mutat be, amely a pedagógusokat is segítheti abban, hogy a valódi tudományos szenzáció ne válhasson a hamis hírek áldozatává (Patel 2018). A kézzelfogható bizonyítékok – például a Holdról visszahozott anyagok – mellett a szerző felhívja a figyelmet a Holdon készült fotókban kételkedők egy, tudatlanságból eredő, alapvetően hibás következtetésére: a többféle szögben észlelhető árnyékokat nem egy stúdiófelvétel lámpái okozzák, hanem a Hold felszínét borító por rigolittartalma. Ez az anyag visszaveri a fényt, éppúgy, mint a tükör, ezért láthatunk különös, a testek helyzetével nem magyarázható árnyékokat a Holdon készült fotókon.

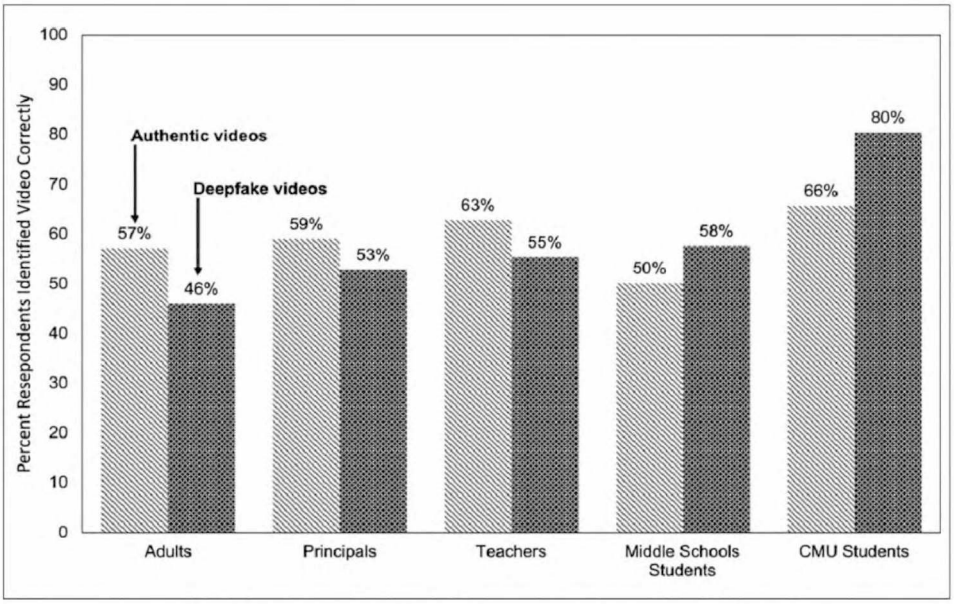


4. ábra. Bal oldalon: az Apollo–14 parancsnoka, Shepard tartja a zászlót a Fra Madura leszállási ponton. Jobb oldalon: az Apollo–15 parancsnoka, Scott tiszteleg a zászlónak Hadley-Apennine-ben. Háttérben: a holdkomp és a holdjáró (forrás: NASA, W4)

A képi tartalmak hitelesítésére ez a cikk is a többféle adatforrás összevetését javasolja: a holdjárók ugyanolyan tájat rögzítettek a Holdon, mint az űrhajósok. A Holdon nem fúj a szél, ezért a legtöbbet vitatott fotó az amerikai zászló volt, amelynek anyaga ráncokat vetve látható. Ezek a ráncok azonban a zászló kibontása és kitűzése közben keletkeztek. Maga a zászló nem lengett. A tanárt az információ hitelességének megítélésében a tudományos magyarázatot közlő hiteles forrás segíti. A diáktól nem várható el, hogy ezeket maga is elolvassa, bíznia kell az oktatójában. A deepfake elleni küzdelemben *az oktatásnak a hiteles forrás szerepét kell betöltenie.*

A deepfake a természettudományok oktatóit különösen nehéz feladat elé állítja: a technológia puszta léte elég ahhoz, hogy az emberek kételkedni kezdjenek a tudományban. Az iskola egyik legnehezebb feladata ma az, hogy segítsen megőrizni a tudományos eredményekbe vetett hitet. Az álhírek a tanári magyarázatnál sokkal jobban „fogyasztható" formában terjednek, és máris jelentős hatást fejtenek ki – elsősorban a felnőtt lakosságra. Az amerikai Challenger Center, a RAND Corporation és a Carnegie Mellon Egyetem *A deepfake és a tudományos ismeretterjesztés (Deepfakes and Scientific Knowledge Dissemination)* című kutatásában különböző életkorú és képzettségű csoportokban vizsgálta a tudományos álhír hatását (Doss et al. 2022). Azt kutatták, képesek-e a résztvevők különbséget tenni a klímaváltozással kapcsolatos deepfake és valódi videók információi között. A témát azért választották, mert valamennyi vizsgált korosztály életére hatással van, sok, könnyen elérhető információ áll rendelkezésre róla, és (különösen az Egyesült Államokban) erősen megosztja az embereket. A kutatásban használt, tíz perc alatt végignézhető videófilm-sorozat első részében négy, egyenként 10-15 másodperces bejátszás volt négy híres környezetvédő (egy neves oceanográfus, egy meteorológus professzor és két klímaaktivista – az egyikük a közismert Greta Thunberg) előadásából. A videók fele valódi filmrészlet volt, a másik fele deepfake, amelyekben a beszélők szájába olyan mondatokat adtak, amelyek szöges ellentétben álltak a valódi véleményükkel. A kísérleti személyek véletlenszerűen kapták meg a valódi és hamisított filmrészleteket. A feladatuk az volt, hogy eldöntsék, melyik a hiteles a négy filmrészlet közül. A vizsgálat második részében kérdőíven kellett számot adniuk saját, a klímaváltozással kapcsolatos tudásukról és vélekedéseikről.

A válaszadók több mint fele helyesen ismert fel néhányat a hamisított filmrészletekből, de hogy hányat, az egyértelműen az életkor és képzettség együttesétől függött: a fiatal egyetemisták (vegyesen társadalom- és természettudomány szakosok) 66%-a azonosította helyesen az autentikus videót, és a deepfake leleplezése sem jelentett nekik problémát: 80%-ban felismerték. A deepfake felismerésében a középiskolások kicsivel jobb eredményt értek el, mint a tanáraik (58, illetve 55%), míg az idősebb, diplomás, de nem tanár végzettségű válaszadók 12 százalékkal rosszabb eredményt értek el. A „nem tudom" és a „nem adok választ" aránya egyik



5. ábra. Deepfake és valódi videók felismerése a klímaváltozás témájában: különböző életkorú amerikaiak válaszai (forrás: Doss et al. 2022: o. n.)

csoportban sem volt kevesebb 20, illetve 30%-nál. A kutatók szerint különösen aggályos, hogy a felnőttek (köztük tanárok) és az iskolavezetők csoportja rosszabb eredményt ért el a hamisított videók azonosításában, mint a középiskolások és az egyetemisták (rövid összefoglaló: Challenger Center 2022; a vizsgálat részletei: Doss et al. 2022). A vizsgálat adatai arra utalnak, hogy *a digitális kompetenciának nagyobb szerepe van a deepfake felismerésében, mint az élettapasztalatnak*. A diplomás felnőttek védtelenebbek a deepfake-tartalmaktól, mivel kevesebb ilyennel találkoztak, és enyhébb bennük a gyanakvás a médiatartalmak hitelességével szemben. Ez az eredmény aggasztó, de hasznos is: megmutatja, milyen irányban érdemes fejleszteni a felnőtt hírfogyasztók tudását. A digitális kompetencia és a vizuális kommunikáció területein képezve őket, a kurrens képalkotó technikák megismertetésével sokat tehetünk azért, hogy a deepfake ellen felvértezzük a társadalmat. A deepfake hasonlóan erős befolyásoló erővel bír a politikai üzeneteknél is, ezért a társadalomismereti képzésen túl az erkölcsi nevelés nézőpontjából is fontos, hogy minden tanár megszerezze ezt a képzettséget.

Egy hamis tény vagy meghamisított esemény akkor válik meggyőződésünkké, ha jórészt megegyezik a világnézetünkkel, ha sokszor halljuk számunkra megbízható információforrásból és hozzánk hasonlóan gondolkodó emberektől. A magas technikai színvonalú, hihető narratívába vagy igazoltan hiteles tudáskörnyezetbe illesztett deepfake-et nagyon nehéz felismerni és elkülöníteni a szereplő

engedélyével, jogszerűen alkalmazott, hasonló technikától, amelyben például egy reklámkampány főszereplőjét „beszéltetik" általa nem ismert idegen nyelveken, hogy a reklám a megcélzott fogyasztók anyanyelvén szóljon. Az egyik legtöbbet hivatkozott kézikönyv, amelyből a tanárok megszerezhetik a deepfake felismeréséhez szükséges médiaértékelési kompetenciát, Rebecca Blankenship (2021) műve, amelynek címe jelzi, milyen komolyan fenyegeti a deepfake az oktatás világát: „Mély hamisítványok, hamis hírek és dezinformáció az online tanítási és tanulási technológiákban". Ez a kötet nemcsak a deepfake elleni védekezésre hív fel, hanem a már beépült hamis tények, adatok, összefüggések azonosításához is segítséget nyújt. A szerző szerint az oktatás számára különösen veszélyesek a *részben manipulált álló- vagy mozgóképek*, amelyeknek egyes részei valóságosak, jól beazonosíthatók, viszont a kép vagy film egésze a valóságos, alapul vett jelenettől részben vagy teljesen eltérő üzenetet hordoz. Az arcelemzésen alapuló deepfake-videók legtöbbje ilyen: a szereplőkkel készült valódi alkotásokat módosítja egy új képelem vagy hangfájl hozzáadásával. Ha ezek hamis tartalma dokumentumként beépül az oktatásba, jelentős kárt okoz, hiszen egy hibásan rögzült, gazdag képi és hangos információt tartalmazó adat „felülírása" pusztán magyarázattal igen nehéz.

A történelem és társadalomismeret oktatásában mindig jelen volt a forráskritika. A képek túlsúlya az információszerzésben, amely a 20. század utolsó évtizedeiben kezdődött, és napjainkban vált uralkodóvá, és ez az új kommunikációs stílus a pedagógus számára *a képi források kritikájának megtanítását* adja feladatul. A hiteles információforrások azonosítása mellett fel kell hívni a figyelmet a képek részleteire, hiszen ezek árulkodók lehetnek, ha a hamisítvány készítője nem elég alapos. A megvilágítás változása egy képen belül, a színárnyalatok logikátlan különbségei egy forma felszínén, az arányok apró hibái mind jelezhetik, hogy a látványt manipulálták. Egy ilyen mű nem tekinthető dokumentumnak addig, míg tartalmát más, hiteles források nem igazolják. Ha pedig nem található ilyen, valószínű, hogy hamisítvánnyal van dolgunk. Remélhetőleg sosem jutunk el abba a korba, amelyről elmondhatjuk: „Azé a történelem, akié a technológia." Ez a kor az oktatás értelmetlenné válását, az igazság relativizálódását jelentené: egy tudománytalan, értékvesztett s ezért embertelen kultúrát.

A jó minőségű hamisítvány azonban emberi szemmel gyakran nem különböztethető meg egy eredeti képtől. A kortárs dokumentumokat remélhetőleg hamarosan megvédhetjük a tartalmukat megváltoztató hamisítványoktól. A Microsoft és a *Defense Advanced Research Projects Agency* (DARPA, W5) párhuzamosan dolgoznak a hamisított, „szintetikus médiának" nevezett hírek gyártóinak leleplezésén. Egy másik jelentős fejlesztés az „általános időpecsét" *(„universal timestamp")*, amely megváltoztathatatlan dátumot rendel a műhöz, ez segíteni fog annak bebizonyításában, hogy a hiteles szöveg vagy kép a hamisítványnál korábban készült el, és eredetinek tekinthető. Nemsokára a deepfake-tartalmak felismerése is digitális eszközökkel történik majd, hiszen a vizuális jelek és képek értelmezésé-

vel foglalkozó kutatóintézetek és fejlesztők sokasága próbál olyan szoftvert létrehozni, amely képes felismerni a manipulált képi tartalmakat. Természetesen, ha a deepfake felismerésére megtanítható egy rendszer, akkor arra is, hogyan manipulálja a deepfake-felismerő alkalmazásokat. A harc az álhírgyárosok és leleplezőik között hosszú lesz, de talán nem reménytelen, hiszen az utóbbiaknak szélesebb tudásbázis, jelentősebb számítástechnikai kapacitás áll rendelkezésére (Nguyen et al. 2022).

A tanárnak nemcsak a deepfake-üzenetek azonosítása a feladata, a szereplők más, hiteles forrásból származó médiamegjelenéseivel összehasonlítva az álinformációk tartalmát. Azt is meg kell értetnie a diákjaival, hogy az ilyen üzenetek készítése élő személyekről jóval több, mint ártatlan tréfálkozás: személyiségi jogokat sért. A deepfake készítése egyre könnyebb és olcsóbb, ezért az oktatásban nem elég csak a kész művek hitelességének ellenőrzéséről beszélni. Nem nagy tárhelyet igénylő alkalmazás például a *DeepFaceLab*, a *FakeApp*, *MyFakeApp* vagy a valós személyek arcait másokkal forgatott jelenetekbe helyező *FaceSwap*. A Facebookon elérhető *MSQRD* applikáció szűrőinek segítségével a diákok könnyen készíthetnek önmagukat híresség között bemutató videót, de ugyanilyen könnyen hamisíthatnak kevésbé valószínűtlen, megtörténtnek tűnő eseményeket is, akár élő videóban. Ha a Reddit a deepfake-forradalom Pilvax kávéháza (Bodnár 2018), elképzelhető, hogy egy diák érdekes időtöltésnek fogja tartani egy valódi forradalom szereplőinek deepfake-esítését. *A fiataloknak tehát tudniuk kell, hogy amit tesznek: hamisítás*, amely éppen olyan súlyos jogi következményekkel járhat, mint ha valakinek az aláírását utánoznák egy szerződésen. A fiatalok médiahasználatát kutatók ezért a deepfake készítését és befogadását is a digitális kockázatok közé sorolják (Guld 2021).

A *médiatudatosság fejlesztésével a közoktatásban* a korábban informatika, most bővített tartalommal digitális kultúra nevű kötelező, 12 éven át oktatott tantárgy és a (sajnos igen kevés osztályban és óraszámban tanított) mozgóképkultúra és médiaismeret, illetve a rajz és vizuális kultúra tantárgy foglalkozik. A felkészítés az információk tartalmán alapuló azonosításra a tartalomhoz kötődő tudományág alapjait oktató valamennyi szaktanár közös feladata (Herzog–Racsko 2018). A médiatudatosság sokat segít a deepfake felismerésében és a hamis információk kiszűrésében. Erre azért is szükség van, mert egy dél-koreai médiapedagógiai kísérlet bebizonyította, hogy a deepfake gyakran hatásosabb, mint a hiteles dokumentum. Kísérletükben 316, 20–59 éves felnőtt vett részt. A résztvevők többsége élénkebb színűnek, meggyőzőbbnek és hitelesebbnek látta a hamisított videót, mint az alapjául szolgáló eredeti filmrészletet. Akik képzést kaptak a deepfake-tartalmak felismerésében, azok viszont – bár közülük is sokaknak tetszett jobban a hamis alkotás – képesek voltak a hiteles tartalmat azonosítani (Hwang et al. 2021).

A deepfake iskolai terjedése tanulási programokkal és biztonsági intézkedésekkel jelentősen lassítható, hatása csökkenthető. A korábban már idézett kanadai

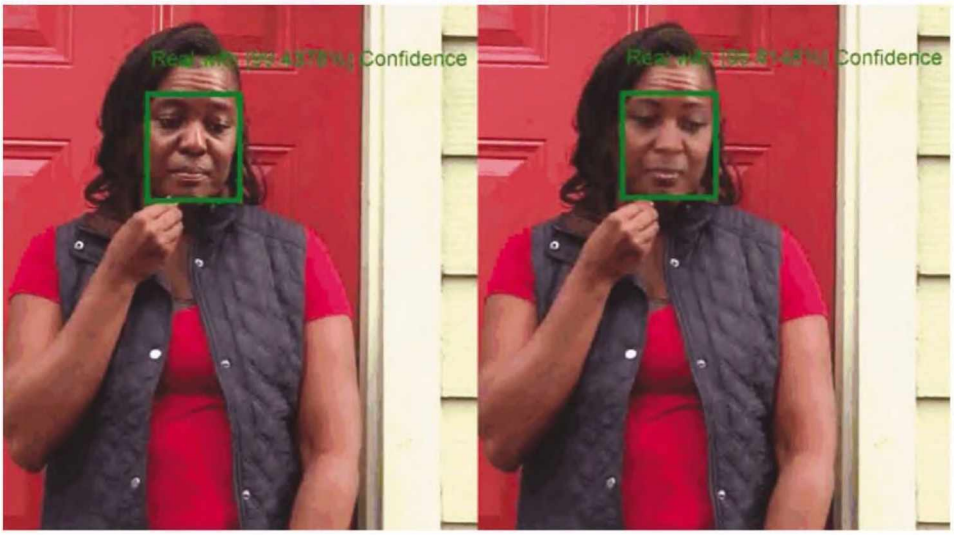
digitális biztonsági szakember, Jordan Foster az iskolavezetőknek szóló online folyóiratban megjelent cikkében (Foster 2021) ezeket javasolja:

- Minden oktatási intézménynek legyen digitális biztonsági terve és csapata: rendszergazda, informatikatanár és a deepfake-témát ismerő más szaktanár.
- Az etikus eszköz- és internethasználatról minél több tantárgy keretében, a tananyaghoz kapcsolva halljanak a tanulók; készüljön etikus digitális eszközhasználati szabályzat a tanároknak és a tanulóknak egyaránt.
- Ellenőrizzék rendszeresen az iskolai számítógépeket és egyéb, közös használatú digitális eszközöket, keressenek deepfake előállítására is alkalmas szoftvereket, figyeljék ezek használatát.
- Az oktatási intézménynek legyen külön intézkedési terve a deepfake-tartalmak ellen. Alaposan vizsgálják ki a panaszokat (például az online zaklatási ügyeket), és gondoskodjanak arról, hogy az áldozatok kapjanak pszichológus segítőt is a történtek feldolgozására.

Mindehhez sok munkaidő és jól képzett szakemberek kellenek. A Bűvösvölgy Médiaértés-oktató Központ (W6) budapesti, soproni és debreceni telephelyein folyamatosan zajlanak a médiakompetencia-fejlesztő pedagógus-továbbképzések és iskolai osztályoknak szóló foglalkozások, az álhírek és hamisított digitális képek témájában is. A Microsoft oktatási programjai évtizedek óta támogatják a digitális innovációt, és segítenek az etikus eszközhasználat elsajátításában is (Kárpáti et al. szerk. 2008). A cég új digitális eszköze, amely képes azonosítani az álhíreket videóanyagok elemzése alapján, több mint két éve elérhető. A szoftver a cég által finanszírozott, és a Princeton Egyetemen, Jacob Schapiro vezetésével lezajlott kutatásra épül, amely sikeresen azonosított 96 álhírkampányt, amellyel 30 országban kísérelték meg 2013–2019 között választások és más, társadalmi alapú, fontos politikai döntések befolyásolását. A deepfake-videókat azonosító eszköz a Microsoft demokráciavédő programjának része (*Defending Democracy Program*, Burt 2020, 6. ábra).

A Yettel ProSuli (korábban Hipersuli) projektje (W7) iskolákat fog össze és támogatja a digitális újításra nyitott pedagógusok munkáját. Ez a program is segít az iskolai digitális visszaélések elleni védekezésben, amelyhez külön aloldalt nyitottak (W8).

Mit tehet még a tanár a hamisított videók ellen? Ha érdekes, hasznos szemléltető anyagnak tartja, amelynek a tartalma tudományosan hiteles, „csak" a fotók alakulnak mozgóképpé, vagy egy, a film felfedezése előtt évszázadokkal élt történeti figura szólal meg a képernyőn, megfontolhatja, hogy felhasználja-e az oktatásban. A deepfake-ről szóló pedagógiai cikkekben ritkák a jó példák, hiszen deepfake-kel tanítani sokak számára elfogadhatatlan félrevezetést jelent. „Deepfake Ady és AI Babits már egymásra tekintget az életre kelt fotón" című, a hatásos oktatási illusztrációk híveinek is fontos cikkében Bodnár Zsolt (2021) egy, furcsa módon



6. ábra. A Microsoft álhírek azonosítására alkalmas szoftverének demo-gifjei (forrás: Burt 2020)

digitális identitásvédelemmel foglalkozó izraeli cég MyHeritage nevű, DNS-alapú családfakutatással foglalkozó oldala egyik szoftverével készült. A *Deep Nostalgia* a régi családi fotók animálása mellett felhasználható történelmi személyiségeket ábrázoló fényképek életre keltésére is. A magyartanár eldöntheti, érdekes vagy riasztó-e a megmozgatott fotó, és ha úgy gondolja, hatásosan egészíti ki a két jeles irodalmár költői munkásságuk szempontjából is fontos barátságáról szóló tanári magyarázatot, akkor nyugodtan felhasználhatja. A kép a deepfake-jelenség sötét oldalának bemutatására is kiválóan alkalmas.

7. ábra. Babits Mihály és Ady Endre fotójából a Deep Nostalgia szoftverrel készült gif (forrás: Bodnár 2021 és Wikimedia Commons / OSZK)

A deepfake képek, más, kevésbé stigmatizáló kifejezéssel, szintetikus képmások *(synthetic media)* iskolai felhasználása mellett érvel Ashish Jaiman (2020) is. Szerinte *a szintetikus képek tartalma dönti el, alkalmasak-e egy személy vagy jelenség hiteles bemutatására*

ra. Könyvében részletesen ismerteti azokat a szintetikus képátalakító és -előállító szoftvereket, amelyeket speciális testi kihívásokkal élők is jól használhatnak saját, kreatív művek előállítására. Ilyenek például a hangelőállító szoftverek, amelyek a laterális szklerózisban szenvedőknek adnak lehetőséget a beszédre, vagy idetartoznak a látássérülteknek készült, a virtuális világokban történő navigációt segítő alkalmazások is. A technika önmagában nem minősíthető, hiszen hiányzó képességeink pótlására is alkalmazhatjuk – véli Jaiman. Az animált fotókon, átalakított filmeken akár hiteles tudományos tényeken alapuló szövegeket is adhatunk a velük összekapcsolható szereplők szájába. Ilyen például John F. Kennedy amerikai elnök 1963-as beszéde, amelyben kísérletet tett a hidegháború lezárására. Ez a fontos történelmi dokumentum korábban csak írásban létezett, de a történelem vagy a politikatudomány tanára immár bemutathatja, hogyan hangzott volna, ha a dallasi merénylő golyója nem oltja ki az elnök életét (W9). A hangelemzésen alapuló beszédrekonstrukció készüléséről és hatásáról szóló dokumentumfilm a 2018-as cannes-i filmfesztiválon díjat nyert, az elnök a filmben megszólaltatott munkatársa pedig döbbenetes hatású alkotásnak tekinti a rekonstrukciót, és hálás érte.

A lényeg tehát a szándék, amely a kép vagy hang (újra)alkotójának kezét vezeti. A pedagógus feladata, hogy eldöntse: bemutatja-e a rekonstruált hangzó anyagot, a festményparafrázist, az animált történelmi fotót. Az oktatás során lehetőség van arra, hogy valódi tények váljanak a tanulók gondolkodásának alapjává, hiszen sokszor, különféle megvilágításban, hiteles helyen és szereplőktől hangzik el az információ. *A legrosszabb, ami történhet az iskolával, ha elveszti hiteleshely-jellegét.* A pedagógusok szerepe a deepfake elleni küzdelemben ezért óriási.

3. MIT TEHET A SZÜLŐ?

Mi kell ahhoz, hogy ne legyen vége a valóságnak? A szülői mediáció és a médiaműveltség fontossága a technopoliszban beszédes című írás (Lerch-Cserei 2021) jelzi, hogy a szülők szerepe a deepfake azonosításában és hatástalanításában alapvető. Az írás a *szülői világértelmezés* szerepét emeli ki. Az „igazság utáni kor" *(post-truth)* kultúrában könnyen hozzáférhető és hatásosan tálalt valós és álinformációk között eligazodni próbáló gyermekek a covid-járvány alatt a szüleik közvetlen közelében éltek rövid megszakításokkal, két éven át. Ez az időszak lehetőséget nyújthatott arra, hogy a gyerek megfigyelje, és közvetlen tapasztalással megtanulja, hogyan viszonyulnak a szülei a világból érkező információkhoz. Olyan oktatási helyzet volt ez, amely az iskolai életben csak a kirándulások, erdei iskolák néhány napos időszakaiban alakulhat ki: a felnőtt élőben demonstrálja, honnan és hogyan kell tájékozódni, kinek érdemes hinni. Még nem tudjuk, mennyire volt hatásos ez az intenzív médianevelés, és mennyi marad a szokásokból a hétköznapokban. A korábbi vizsgálatok azt mutatják,

a szülők tudatában vannak a média hatásának önmagukra és a gyerekeikre, és sokféle módon próbálják szabályozni ezt a hatást.

A *szülők médiahasználattal kapcsolatos magatartásának* vizsgálatakor ezeket a nevelési típusokat azonosították a kutatók:

1. együttes használat: a szülő és a gyerek együtt használja a médiát, például közösen néz televíziót, vagy együtt játszik az interneten;
2. aktív mediáció: a szülő beszélget a gyerekével a médiatartalmakról;
3. korlátozó mediáció: a szülő akadályozza gyereke médiahasználatát;
4. megfigyelés: a gyerek médiafogyasztását utólag ellenőrzi a szülő;
5. technikai korlátozás: appok segítségével korlátozzák a gyerek médiatevékenységét (Aczél–Andok–Bokor 2015: 173).

A deepfake szempontjából az együttes használat és az aktív mediáció a legfontosabb, hiszen a szülőnek ilyenkor van alkalma arra, hogy bemutassa, milyen a kritikus médiafogyasztói magatartás. A szülő a médiatudatosság fejlesztésében alapvető felhasználói tudást is ilyenkor, a gyakorlatban adhatja át, érzékeltetve, hogy minden médium a maga műfajában sajátos világot alkot, és mesterséges valóságokat mutat be. Ez a médiavalóság pedig saját valóságérzékelésünket is befolyásolja. Különösen igaz ez a deepfake-re, amely már régen túllépett a viccen, és nézeteink befolyásolására tör.

„Az emberek egyfelől beleláradhatnak abba, hogy mindenhol manipulált üzenetekkel, hírekkel, tartalmakkal szembesülnek, ezért Ovadya szóhasználatával »valóságapátia« lesz úrrá rajtuk, így az addig valósnak vélt hírforrásokba, digitális eszközökön keresztül tartott kapcsolataikba vetett bizalmuk elvesztésével elfordulnak a megszokott platformoktól, ezzel viszont a demokrácia működéséhez szükséges általános tájékozottságnak is lőttek" (Bodnár 2018: o. n.).

A hírfogyasztó mint szülő nem eshet „valóságapátiába", nem dönthet úgy, hogy „digitális bennszülött" gyermekére bízza a hírekben rejlő igazság megítélését. Élettapasztalatát kell segítségül hívnia, illetve ugyanazokat az eszközöket, amelyeket a tanár is használ: a hírforrás megítélését és a hír ellenőrzését más, bizonyítottan hiteles forrásból. Nem minden hírnél kell ilyen alaposan eljárnunk, de mindig ellenőriznünk kell, ha a gyermekünk olyan hírre bukkan, amely alapvető értékeinket támadja, amely tudománytalan vagy más szempontból káros. Az életünk alapértékei ellen támadó deepfake olyan, mint a földből előbukkanó, gránátnak kinéző fémdarab: azt sem hajítanánk félre alapos vizsgálat nélkül. Szakembert hívnánk, hogy hatástalanítsa. Ez a szakember a pedagógus, társunk a deepfake elleni küzdelemben.

SZAKIRODALOM

Aczél Petra – Andok Mónika – Bokor Tamás 2015: *Műveljük a médiát!* Budapest: Wolters Kluwer.

Aczél Petra 2017: Az álhír. Kommentár a jelenség értelmezéséhez. *Századvég*, 2. 5–26.

Barber, Gregory 2019: Deepfakes Are Getting Better, But They're Still Easy to Spot. *Wired*, május 26. https://www.wired.com/story/deepfakes-getting-better-theyre-easy-spot/

Bencze Áron 2022: A deepfake-videók veszélyei. *Innotéka*, szeptember 7. https://www.innoteka.hu/cikk/a_deepfake_videok_veszelyei.2581.html

Blankenship, Rebecca 2021: *Deep Fakes, Fake News, and Misinformation in Online Teaching and Learning Technologies*. Hershey: IGI Global.

Bodnár Zsolt 2018: A deepfake éve, a valóság vége. *Qubit*, június 6. https://qubit.hu/2018/06/06/2018-a-deepfake-eve-a-valosag-vege

Bodnár Zsolt 2021: Deepfake Ady és AI Babits már egymásra tekintget az életre kelt fotón" *Qubit*, március 1. https://qubit.hu/2021/03/01/deepfake-ady-es-ai-babits-mar-egymasra-tekintget-az-eletre-kelt-foton?fbclid=iwar0ficnz7m8cmh9fptyaxywnwrucbbuy_-6c0si6jsyd2ik-biup1mmjp9e

Bodnár Zsolt 2022: A diákok már azzal vágnak fel, hogy mesterséges intelligenciával íratják a házi dolgozatokat. *Qubit*, szeptember 28. https://qubit.hu/2022/09/28/a-diakok-mar-azzal-vagnak-fel-hogy-mesterseges-intelligenciaval-iratjak-a-hazi-dolgozatokat

Burt, Tom 2020: New Steps to Combat Disinformation. *Microsoft Blog*, szeptember 1. https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/

Challenger Centre 2022: *Deepfake Science: How Vulnerable Are Educators and Learners*? Washington DC: Challenger Centre for Space Science Education. https://www.challenger.org/2022/08/08/deepfake-science-how-vulnerable-are-educators-and-learners/

Cybertalk 2021: Deepfake technology, are we prepared? *Cybertalk*, október 21. https://www.cybertalk.org/wp-content/uploads/2021/10/shutterstock_1498889558.jpg

De Guise, Lucien 2022: The hidden Catholicism of Vermeer. *Aleteia*, január 19. https://aleteia.org/2022/01/19/the-hidden-catholicism-of-vermeer/

Dombrovski, Eileen 2018: „Deepfakes" and TOK: more trouble ahead for critical thinking? *Oxford Education Blog*, augusztus 27. https://educationblog.oup.com/theory-of-knowledge/deepfakes-and-tok-more-trouble-ahead-for-critical-thinking

Doss, Christopher – Monschein, Jared – Shu, Dule – Wolfson, Tal – Pardee, Frederick – Fitton-Kane Valerie, Kopecky, Denise – Bush, Lance, Tucker, Conrad 2022: *Deepfakes and Scientific Knowledge Dissemination*. Washington DC: Challenger Centre for Space Science Education. https://assets.researchsquare.com/files/rs-1408525/v1/120022c7-acac-45a6-b4c4-e0aea2147a02.pdf?c=1646939243

Foster, Jordan 2021: Is Your School at Risk of a Deepfake Attack? *School Governance*, november 24. https://www.schoolgovernance.net.au/news/is-your-school-at-risk-of-a-deepfake-attack

Guld Ádám 2021: *A Z generáció médiahasználata: Jelenségek, hatások, kockázatok*. Budapest: Libri.

Herzog Csilla – Racsko Réka 2018: A médiatudatosság fejlesztésének lehetőségei a digitális átállás korában. In: Nádasi András (szerk.): *Agria Média 2017. Magyar és angol*

nyelvű konferenciakötet. Eger: EKE Líceum Kiadó. 27–33. DOI: 10.17048/AM.2018.27 http://publikacio.uni-eszterhazy.hu/id/eprint/2331

Hwang, Yoori – Youn Ryu, Ji – Jeong, Se-Hoon 2021: Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education. *Cyberpsychology, Behavior, and Social Networking*, 24/3: 188–193. DOI: 10.1089/cyber.2020.0174

Jaiman, Ashish 2020: *Deepfakes & Synthetic Media: Humanity at the Edge of an Uncanny Valley*. e-könyv, szerzői kiadás. ISBN: 13. 979-8843038922.

Kárpáti Andrea – Molnár Gyöngyvér – Tóth Péter – Főző Attila szerk. 2008: *A 21. század iskolája*. Budapest: Nemzeti Tankönyvkiadó. https://mek.oszk.hu/20400/20407/20407.pdf

Kárpáti Andrea – Nagy Angelika 2019: Digitális kreativitás – a vizuális és informatikai kultúra szinergiája. *Iskolakultúra*, 29/4–5: 86–98. https://doi.org/10.14232/ISKKULT.2019.4-5.86

Kovach, Bill – Rosenstiel, Tom 2011: *Blur. How to Know What's True in the Age of Information Overload.* New York: Bloomsbury.

Landau, Shira 2021: Teaching Children About Deepfake Technologies. *E-Learning Industry,* november 27. https://elearningindustry.com/teaching-children-about-deepfake-technologies

Lerch-Cserei Judit 2021: Mi kell ahhoz, hogy ne legyen vége a valóságnak? A szülői mediáció és a médiaműveltség fontossága a technopoliszban. *Korunk*, 8: 73–78. https://epa.oszk.hu/00400/00458/00675/pdf/EPA00458_korunk_2021_02_073-078.pdf

Lopez, Jonathan 2009: *The Man Who Made Vermeers: Unvarnishing the Legend of Master Forger Han Van Meegeren.* Boston: Mariner Books

Nguyen, Thanh Thi – Quoc Viet Hung Nguyenb, Dung Tien Nguyena, Duc Thanh Nguyena, Thien Huynh-Thec, Saeid Nahavandid, Thanh Tam Nguyene, Quoc-Viet Phamf, Cuong M. Nguyen 2022: Deep Learning for Deepfakes Creation and Detection: A Survey. *Computer Vision and Image Understanding* 223. DOI: https://doi.org/10.1016/j.cviu.2022.103525

Patel, Neel V. 2018: 7 easy ways you can tell for yourself that the moon landing really happened – The proof is out there. *Popular Science,* december 11. https://www.popsci.com/proof-moon-landing-not-fake/

SWW Education 2021: Deepfake teachers and technology: the future of K-12 public education? *SWW Education*, szeptember 18. https://swweducation.org/deepfake-teachers-technology-the-future-of-k-12-public-education/

Szerző nélkül. 2011: Meegeren: a hamisító, aki átverte a világot. *Múlt-kor,* szeptember 19. https://mult-kor.hu/cikk.php?id=34394

Szűcs Zoltán 2020: Az online lét diszkrét digitális bája. A post-truth korához vezető út. *Létünk*, 4: 11–25.

Veszelszki Ágnes 2017: Az álhírek extra- és intralingvális jellemzői. *Századvég* 84: 51–82.

Veszelszki Ágnes 2021: deepFAKEnews: Az információmanipuláció új módszerei. In: Balázs László (szerk.): *Digitális kommunikáció és tudatosság.* Budapest: Hungarovox Kiadó. 93–105.

Veszelszki Ágnes – Horváth Evelin – Mezriczky Marcell 2022: Manipulált képi és deepfake-tartalmak felismerésének lehetőségei. In: Hulyák-Tomesz Tímea (szerk.): *A digitális oktatás tapasztalatai a kommunikációs készségfejlesztésben.* Budapest: Kommunikációs Nevelésért Egyesület. 85–99.

Yadav, Digvijay – Salmani, Sakina 2019: Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network. In: *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*. Madurai: IEEE. DOI: 10.1109/ICCS45141.2019.9065881

FORRÁSOK

Rae, Jack et al. 2021: Scaling Language Models: Methods, Analysis & Insights from Training Gopher. *DeepMind Papers,* december 8. https://storage.googleapis.com/deepmind-media/research/language-research/Training%20Gopher.pdf
W1 = Content Authenticity Initiative. https://contentauthenticity.org
W2 = Han van Meegeren. Wikipédia. https://hu.wikipedia.org/wiki/Han_van_Meegeren
W3 = https://www.youtube.com/watch?v=6PndwgJuF3g
W4 = NASA, https://www.nasa.gov/feature/flag-day-flying-high-the-stars-and-stripes-in-space
W5 = DARPA, https://www.darpa.mil/
W6 = Bűvösvölgy Médiaértés-oktató Központ, https://buvosvolgy.hu/
W7 = Yettel ProSuli, https://prosuli.hu/hipersuli/
W8 = https://prosuli.hu/tudatosnet/miert-lehet-veszelyes-a-deep-fake-jelenseg/
W9 = The Times 2018: JFK Unsilenced. https://www.youtube.com/watch?v=wZF59wIIBLI

Szerzőink

Prof. dr. Aczél Petra, PhD. Nyelvész, kommunikációkutató, retorikus, a Moholy-Nagy Művészeti Egyetem egyetemi tanára. Az NMHH Médiatudományi Intézetében kutatásvezető. A Magyar Tudományos Akadémia Kommunikáció és Médiatudományi Bizottságának alelnöke. Az MTVA mellett működő nyelvi tanácsadó Montágh Testület elnöke. Oktatott a Budapesti Corvinus Egyetemen, a Pázmány Péter Katolikus Egyetemen, az Eszterházy Károly Főiskolán, az Eötvös Loránd Tudományegyetemen, illetve az Egyesült Államokban a University of Richmondon. Számos kutatási projekt vezetője (különösen médiaértés, tudománykommunikáció, időskép, művészeti kommunikáció, társadalmi jövőképesség témákban). Fő kutatási területe a retorika és az újmédia-kommunikáció.

Dr. Eszteri Dániel, PhD. Jogász, a Nemzeti Adatvédelmi és Információszabadság Hatóság osztályvezetője. Az Eötvös Loránd Tudományegyetem Jogi Továbbképző Intézet és a Nemzeti Közszolgálati Egyetem megbízott oktatója adatvédelmi jogból. 2015-ben PhD-fokozatot szerzett a Pécsi Tudományegyetemen a virtuális tulajdonról írt disszertációjával.

Dr. Gocsál Ákos, PhD. A Pécsi Tudományegyetem Művészeti Karának egyetemi adjunktusa, a Zeneművészeti Intézet igazgatóhelyettese, illetve az ELKH Nyelvtudományi Kutatóközpont tudományos munkatársa. Fonetikával 1998 óta foglalkozik. Kutatásaiban azt vizsgálja, hogy a beszélő személy egyes tulajdonságai – neme, életkora, testalkati paraméterei, személyisége stb. – a beszéd hangzásában hogyan tükröződnek, illetve a hallgató a beszéd hangzása alapján milyen benyomást alakít ki a beszélőről. Publikációi más területeken is jelennek meg. Érdeklődése kiterjed a zenetanulás transzferhatásaira, különösen a beszédkommunikációs jelenségek vizsgálatára zenészek és zenei képzettség nélküli beszélőknél. További írásai a projektpedagógiával, médiapedagógiával, régi filmekkel, filmhíradókkal, oktatástechnológiai témákkal foglalkoznak. Nyolc külföldi egyetemen tanított vendégoktatóként, emellett számos tudományos és ismeretterjesztő előadást tartott hazai és nemzetközi szakmai rendezvényeken.

Dr. Guld Ádám, PhD. Médiakutató, kommunikációs szakember, tanácsadó, a PTE BTK Kommunikáció- és Médiatudományi Tanszék habilitált egyetemi docense, a Magyar Tudományos Akadémia Bolyai János Kutatási Ösztöndíj ösztöndíjasa. Alapító tagja és titkára a Neumann János Számítógép-tudományi Társaság eHétköznapok Szakosztályának, illetve a médiatudományi oktatással, kutatással és tehetséggondozással foglalkozó Médianegyed Egyesületnek. Az elmúlt években három önálló kötetet publikált, ezek közül a legfrissebb 2022 májusában jelent meg *A Z generáció médiahasználata: jelenségek, hatások, kockázatok* címmel.

Horváth Evelin. A Budapesti Corvinus Egyetem Szociológia és Kommunikációtudomány Doktori Iskolájának doktorandusza. Tíz éve foglalkozik portréfotózással, e hivatásához kötődően elsődleges kutatási területét a vizuális kommunikáció újszerű megnyilvánulási formái jelentik: a digitális képmanipuláció, a CGI-technológia, valamint a deepfake társadalmi hatásai, marketinglehetőségei.

Prof. dr. Kárpáti Andrea, PhD. A Budapesti Corvinus Egyetem Marketing- és Kommunikációtudományi Intézetének egyetemi tanára, az MTA doktora, a Vizuális Kultúra Kutatócsoport vezetője. A European Visual Literacy Network (Európai Vizuális Műveltség Hálózat) kutatói közösség alapító tagjaként részt vett a Common European Framework of Visual Literacy (Európai Vizuális Kompetencia Referenciakeret) kialakításában. Kutatási területei: a vizuális képességek fejlődésének vizsgálata (gyermekrajz és az ifjúsági szubkultúrák képi nyelve) hagyományos és digitális médiumokban, a vizuális tehetség felismerése és gondozása, kompetenciaalapú képességmérés. 2022-ben is a HORIZON 2020-as kutatási programban hét ország részvételével zajló, *Acting on the Margins: Arts as Social Sculpture* (AMASS, „Művészetpedagógia a társadalom perifériájára sodródottakért") című kutatás magyar munkacsoportjának vezetője.

Keleti Arthur. Kibertitok-jövőkutató, IT-biztonsági stratéga, előadó, író, filmproducer. A 2021-ben nagy sikerrel zajlott, többnapos, virtuális tévéprogrammal is jelentkező Informatikai Biztonság Napja (ITBN) konferencia ötletgazdája és főszervezője, amelynek keretében már 18 éve hívja össze a magyar biztonsági piac szereplőit közös gondolkodásra. Az Önkéntes Kibervédelmi Összefogás (KIBEV) alapítója és elnöke. A *The Imperfect Secret* című könyv szerzője, a kibertérben megjelenő személyes és üzleti titkok biztonságának, a kiberbiztonság jövőjének kutatója. Kutatóként nemzetközi és hazai rendezvények előadója, televízió- és rádióműsorok állandó szereplője, elsősorban a kiberbiztonság jövőjével, a digitalizáció, a mesterséges intelligencia és az emberi és szervezeti titkok digitalizálásának társadalmi és technológiai hatásaival foglalkozik. A *Sight: Extended* című amerikai film executive producere, amely a jövőben játszódik, ahol intelligens kontaktlencsékkel helyettesítik az okostelefonokat. 1999 óta dolgozik az ICON, a KFKI, a T-Systems Magyarország, majd a Magyar Telekom biztonsági csapatával.

Dr. Krasznay Csaba, PhD. A Nemzeti Közszolgálati Egyetem docense, kutatási témája a kiberbiztonság. Jelenleg az egyetem Kiberbiztonsági Kutatóintézetének intézetvezetője, emellett a kiberbiztonsági mesterképzés és az elektronikus információbiztonsági vezető szakirányú továbbképzés szakfelelőse. Rendszeres vendégoktató az ELTE-n és az Óbudai Egyetemen, tudományos munkatárs a Tallinni Műszaki Egyetemen. 2003-ban szerezte meg diplomáját a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Kar villamosmérnöki szakán, majd PhD-ját az NKE-n 2012-ben katonai műszaki tudományok területén. 2011-ben az év útmutató biztonsági szakemberének választották. Az Önkéntes Kibervédelmi Összefogás jelenlegi, a Magyary E-közigazgatástudományi Egyesületnek, az ISACA magyar tagozatának és a Magyar Elektronikus Aláírás Szövetségnek korábbi elnökségi tagja. Emellett tagja a Magyar Hadtudományi Társaságnak és a Hírközlési és Informatikai Tudományos Egyesületnek. Felsőoktatási tevékenysége mellett folyamatosan dolgozik piaci közegben is, az OXO Cybersecurity Lab kiberbiztonsági startup inkubátor alapítója. Rendszeres előadója hazai és nemzetközi konferenciáknak, valamint kiberbiztonsági témákban az egyik legtöbbet foglalkoztatott médiaszakértő Magyarországon. Önkéntes munkája során a Nemzetközi Gyermekmentő Szolgálat szervezésében általános és középiskolákban tart órákat a gyermekeknek és a szülőknek a biztonságos internethasználatról.

Dr. Lendvai Gergely Ferenc. Jogász, a Nemzeti Média- és Hírközlési Hatóság jogi szakértője és a Pázmány Péter Katolikus Egyetem Jog- és Államtudományi Kara Doktori Iskolájának PhD-hallgatója. Jogi diplomáját az Eötvös Loránd Tudományegyetemen, jogi mesterdiplomáját pedig az Université Paris-Panthéon-Assas (Párizs II) egyetemen szerezte. Főbb kutatási területei a platformszabályozás, az online gyűlöletbeszéd, a szólásszabadság és a véleménynyilvánítás szabadsága az online térben, a Digital Services Act interpretációi és implementációja, illetve a jogi „appifikáció" és az óriásplatformok (VLOP) „kvázi-bíráskodása". A szerző külsős oktató a Budapesti Corvinus Egyetemen és a Károli Gáspár Református Egyetem Állam- és Jogtudományi Karán, továbbá a Monroe E. Price nemzetközi médiajogi perbeszédversenyen az ELTE ÁJK csapatának felkészítő tanára.

Mezriczky Marcell. Kommunikációs szakértő, deepfake-kutató. A Budapesti Corvinus Egyetem Szociológia és Kommunikációtudomány Doktori Iskolájának hallgatója, a CGI-Deepfake Kutatócsoport tagja. Fő kutatási témái: deepfake, digitális videómanipuláció, mesterséges intelligencia. Tanulmányai mellett médiatervezőként dolgozik.

Dr. Miklós Gellért. Jogász, infokommunikációs szakjogász. Jelenleg az Óbudai Egyetem Biztonságtudományi Doktori Iskolájának doktorandusz hallgatója. Tanulmányai és kutatása középpontjában a kiberbiztonság, adatbiztonság és adatvédelem hazai és nemzetközi szabályozása áll. Emellett egy nemzetközi távközlési vállalat jogszabályi megfeleléssel foglalkozó munkatársa, szakterülete a dolgok internete (IoT). Napi szinten foglalkozik e témakörben releváns jogszabályok, jogszabálytervezetek értékelésével, a különböző termékek, szolgáltatások jogszabályi megfelelőségének vizsgálatával.

Szabados Levente. Eredeti végzettségét tekintve programozó, ám a tudat és a folyamatai iránti elkötelezettsége előbb a buddhista teológusi mesterképzése során, majd a kognitív tudomány területén végzett kutatásaihoz vezették. 2010-ben az Információs Társadalomért Alapítvány kutatócsoportjának tagjaként a tudásleképezés és szakértői rendszerek területén dolgozott, majd kutatásvezetőként áttért a pénzügyi modellezés és a gépi tanulás területére. Ezt követően egy nyelvtechnológiai területen működő startup technológiai igazgatójaként szerzett tapasztalatot a legmodernebb gépi nyelvfeldolgozási megoldások fejlesztésében és üzleti alkalmazásában. Jelenleg a Frankfurt School of Finance and Management mellett az SRH Heidelberg és az Aivancity Paris vendégprofesszora (gépi tanulás és deep learning témakörben), továbbá a Neuron Solutions gépi tanulással foglalkozó cég vezető tanácsadója. Elérhetőség: levente.szabados@neuronsolutions.hu.

Tari Annamária. Klinikai szakpszichológus, pszichoterapeuta, pszichoanalitikus. Rendszeresen foglalkozik a fogyasztói társadalom hatásaival, a média és az emberi tényezők összefüggéseivel, a társadalmi változások egyénekre ható vonásaival. Az információtechnológia fejlődésével együtt járó lélektani jellemzők, generációs különbségek és az online élet személyiségre ható változásai a fő kutatási területe. 2010-ben megjelent könyve az Y generációt helyezi a középpontba, 2011-ben írta meg a *Z generáció* című kötetét, amely az információs korban élő (kis)kamaszok magatartásváltozásaival és a felnőtt társadalom működésével foglalkozik. 2013-ban a *Ki a fontos: Én vagy Én?* című kötete a magánéletünkben is mélyülő társadalmi nárcizmust, illetve az online és offline tér kapcsolatát, hatásait elemzi az X generáció szempontjából. A 2015-ben megjelent *Rád találni* című kötete az Index Dívány Kettőslátás rovatában megjelent, Horváth Gergellyel közösen írt publicisztikáiból válogat. A *Generációk online* című kötetében (2015) az online tér és az emberi kapcsolatok változásait vizsgálja. A *Bátor generációk* című könyve (2017) a 21. századi szorongásokat taglalja, az *Online illúziók – offline valóság* (2019) című kötetében a valóság megélése, az érzelmek fejlődése és a közösségi média összefüggéseivel foglalkozik. 2021-ben jelent meg *Pillanatnyi boldogságok* című kötete, amelyben egy saját kutatás adatain keresztül a digitalizáció és a személyiség érzelmeinek változása kapcsolatát elemzi, különös tekintettel a kifáradás, kiégés, halogatás és depresszív érzések megjelenését taglalva.

Dr. Veszelszki Ágnes, PhD. Nyelvész, közgazdász, kommunikációkutató. A Nemzeti Közszolgálati Egyetem ÁNTK Digitális Média és Kommunikáció Tanszékén tanszékvezető egyetemi docens. A Nemzeti Média- és Hírközlési Hatóság Médiatudományi Intézetében kutatásvezető-helyettes. A *Filológia.hu* MTA-folyóirat főszerkesztője. Kutatási területei: az audiovizuális manipuláció formái, a deepfake; a mesterséges intelligencia kommunikációs hatásai; tudománykommunikáció. E-mail: veszelszki.agnes@mtmi.hu. Honlap: www.veszelszki.hu.

Abstracts in English

Veszelszki, Ágnes: Deepfake: doubt in disbelief
The chapter explores three dilemmas related to deepfake and the underlying artificial intelligence that recognises and creates images. The first is the towards/away scale, with the negative consequences of automatic image and person recognition (such as the associated social evaluation system and biased algorithm) on one side and the positive earnings (especially social security) on the other. The second dilemma, which can be described with a pro and con, lists the advantages of using deepfake (in business, video technology, entertainment), while weighing up the disadvantages and dangers (including: changing the idealised body image, new forms of cybercrime, redefining the concept of authenticity). Specifically linked to the latter, to the question of authenticity, is the third dilemma, which presents the politics of visual evidence and the paradox of the doubt of doubt.

Keywords: image generating and image recognition algorithm, video manipulation, text mining, credibility

Aczél, Petra: Deepfake as deception – Interactive cooperation in the creation of lies
A media-deepfake is a simulated version of a person or event, it has referentiality but is without its own reality, and as such is deceptive. Audio-visual content produced through deep learning programming can therefore be interpreted as a kind of deception, which is very difficult for the recipient to detect. nevertheless, this is typically not due to the perfect execution of the deception itself, but rather to the fact that human communication is fundamentally characterised by the cooperative intentions of the parties. It is this intention that allows the recipient, by consenting to the interaction, to actually endorse the lie to unfold. By presenting the communication-theoretical aspects of deception, this paper aims to shed light on the background of the reception of deepfakes and the acceptance of deception. Starting from a message-focussed definition of deepfake, it first takes into account information-centred and then interaction-centred theories of deception in order to describe deepfake as deception.

Keywords: deception theory, information manipulation, interaction, deception, cooperation

Mezriczky, Marcell: Do not believe their eyes! An analysis of online press representation of deepfake between 2018 and 2022

People perceive AI-generated faces as more authentic than real ones. To help distinguish manipulation and reality, media offers interpretative frameworks for the users. By doing so, the context in which deepfake is embedded is becoming more important. The aim of my research is to show how the online press connects the technology with different topics and how its representation changes from the first Hungarian media coverage in 2018 to 2022. My research has been conducted with content analysis, examining 882 online publications with the keyword 'deepfake'. Results show that the number of these articles is increasing year by year, with only 25 units in 2018 but already 324 units in 2022. The most frequent categories were "Cybersecurity", "Entertaining" and "Crime, fake news". If positive uses are low-represented in the media, and the "threat" narrative is overwhelmingly dominant, it may also affect how the users relate to the opportunities of the technology.

Keywords: deep learning, artificial intelligence, press representation, content analysis

Keleti, Arthur: Not everything is what it wants to look like – The present and future of deepfake and authenticity

The chapter provides a basis for understanding the deepfake phenomenon by exploring the impact of the relationship between trust, technology and perception in the digital space on the perceived truth and the resulting authenticity in the receiving human. It explores the informational, technological and human contexts of authenticity and describes the technical and human relations of deepfake. It discusses the cybersecurity implications of digital content and its reception. It provides an overview of the relationship between artificial intelligence, video and voice faking and security, and the technological steps and opportunities leading to the emergence of video, voice manipulation and deepfake. Finally, it reviews the potential for misuse of AI and deepfake, some of the incidents and the recommendations, methods and solutions for protection.

Keywords: cybersecurity, trust, authenticity, AI, protection

Szabados, Levente: A deep dive into the world of "deep forgery". The present and (near) future of deepfake technology

We cannot believe our eyes. Deepfake technology, a scientific breakthrough since 2012, has been advancing since 2017 as a special subset of deep neural network learning (deep learning), which produces images, sounds and moving images of a quality that can deceive human perception. The aim of this chapter is to provide a more accurate picture of the scientific and software development dynamics, social awareness and use of deepfake technology using quantitative and qualitative tools, and to estimate its near-term dynamics.

Keywords: technology, development models, quantitative analysis

Krasznay, Csaba: Cybersecurity implications of deepfake technology

Artificial intelligence technologies have been challenging cybersecurity professionals for years. Among these, deepfake is one implementation that has been the subject of numerous studies on its role and threats, but where meaningful technological defence has not yet been widely deployed. Meanwhile, in the areas of cybercrime, cyber warfare, cyber espionage and hacktivism, there are a number of reports suggesting that this solution is being actively used by the attackers. A particular focus is the Russian-Ukrainian war, where deepfakes have been used by both sides right from the beginning of the clashes. The aim of this paper is to review, through case studies, the real risks of deepfake in 2022–2023 and to propose how the defence of cyberspace against this specific risk can be improved.

Keywords: cyber warfare, cybercrime, cyber espionage, hacktivism

Lendvai, Gergely Ferenc: Deepfake in the light of freedom of expression – reflections from a legal perspective

The chapter aims to outline the legal regulation of deepfake with a particular focus on the freedom of expression. After a conceptual basis, the descripton of the relevant US, European and alternative regulations and proposed solutions follows, in the context of unlawful deepfake content. The chapter pays particular attention to the relationship between the exercise of freedom of expression and deepfake technology and its legal assessment. In addition to national and international regulations, the study will also draw on a number of cases considered to be "key cases".

Keywords: freedom of expression, AI regulation, case law

Eszteri, Dániel: Data protection evaluation of deepfake technology in the context of the GDPR

The aim of the chapter is to assess deepfake technology in terms of the fundamental right to the protection of personal data. I analyse data protection problems in the use of the technology on the basis of the relevant provisions of the European Union's General Data Protection Regulation (GDPR). The chapter begins with the historical antecedents of personality theft and, in this context, the manipulation of reality. After this, the background of the technology in general will be presented, primarily from the point of view of the personal data processing operations that each part of the process (training the software, preparing the deepfake, and then using it) entails. In connection with these, the responsibility of the data controller is also separated. Finally, the chapter concludes by outlining the current enforcement options and the latest legislative directions to address the phenomenon. According to the findings, the risks posed to the privacy of data subjects associated with the use of this technology are currently not effectively managed in most cases, due to the extremely difficult identification of data controllers producing the illegal content. However, the current directions of EU legislation include for-

ward-looking measures, the practical implementation of which, however, is also challenging.

Keywords: personal data, data protection, GDPR, machine learning, artificial intelligence, identity theft

Miklós, Gellért: Regulation of deepfake content in European Union law

The European Union identified fake news and illegal deepfake content as one of the biggest challenges to the democracies of the member states and to its own existence. Due to the explosive development of artificial intelligence, new technologies have been developed, which can be used to create audio-visual content that is more lifelike than ever before. The EU has therefore adopted new legislation and amended old legislation to create a legislative framework that promotes the harmonization of the single internal market and the protection of fundamental rights and the development of an innovation-based data economy. In order to more effectively protect fundamental rights and social interests, the range of obligations of online platform service providers has been broadened and strict transparency measures have been established. Different laws regulate different stages and different actors in the creation and distribution of illegal deepfake content.

Keywords: artificial intelligence, European Union, online platform regulation, GDPR

Gocsál, Ákos: Manipulated speech in researching social perception

By the application of deepfake technologies, realistic-looking virtual persons are generated. Voice quality is of importance, not only with real persons, but also with virtual characters, because the listener's acoustic experience may determine his/her attitude towards the speaker or may even establish decisions. It is the researcher's task to find the acoustic parameters of speech that play a role in forming this kind of impressions about the speaker. In the present study, 51 university students were played male and female utterances, nine by each gender. Using 7-point scales, subjects were asked to indicate how pleasant they found the heard speech sample, and they also indicated their decisions in four imaginary situations. Only one male and one female speech sample was natural, while pitch and speech rate manipulations were administered in different combinations on the rest of the utterances. The results suggest a general tendency that lower pitch and faster speech evoked more positive attitudes in listeners, while raised pitch and slower speech were associated with less positive attitudes. The halo effect was also demonstrated, the more pleasant sounding a speech sample was to the listeners, the more accepting they were to the virtual speaker in the different imaginary situations.

Keywords: speech, manipulation, impression, naive description, halo effect

Guld, Ádám: Deepfake and CGI technology in the service of influencer marketing: how digital characters are reshaping the operation of the recognition industry

In the contemporary media environment, we must take the warning not to believe our eyes more and more seriously. Since the end of the 2010s, digitally created influencers and deepfake applications have been reshaping the world of fame and fandom, as digitally created or digitally manipulated influencers are increasingly popular. CGI (computer-generated images) and deepfake (images created with the help of AI and deep learning) characters are not only innocent toys anymore, but also solutions with serious cultural and economic impact. In the study below, I will show what technological solutions work behind digitally created characters, what types of characters we can distinguish, and how the audience reacts to them. Finally, through five short case studies, I interpret the economic potential of the solution and how it is received by consumers.

Keywords: CGI influencers, influencer marketing, online celebrities, effects and reception

Horváth, Evelin: Can beauty be faked? Exploring the relationship between deepfake and the ideal of beauty

Deepfake technology became a worldwide phenomenon a few years ago due to an application being shared on social media that automatically replaced the faces of characters in a video content with other people's faces, while still keeping the video lifelike. Since then, artificial intelligence-based learning software has evolved considerably, making a wide range of image, sound and video manipulation applications available to everyday media users. However, hyper realistic manipulated contents can also deceive viewers. For example, automatically retouched, idealised portraits convey an unattainable ideal of beauty to the viewers, thus negatively affecting people's own body-image. Although there are many criticisms of the use of deepfake, the technology can also be used to draw viewers' attention to important social messages.

Keywords: beauty; ideal of beauty; artificial intelligence; image manipulation; self-image

Tari, Annamária: The psychology of manipulated images and videos: expanding reality, or turning illusions into reality?

The social narcissism of the 21st century and the strong presence of online space indicates changes in the emotional functions of one's personality. Typical behaviour of the 21st century shows significant coherence with the speed of the digital age, the need of immediate emotional gratification and emotional incontinence resulting in adult anger management that might be evaluated as infantile. It is clear that human emotions are not unaffected by information technology and the

megatrends of acceleration. So when it comes to controlling our emotional functions, it is crucial how loose reins we run on. Now we must learn how to use the tools of this new world carefully and wisely in order to prevent the emotional qualities of our human relations from a previously unknown effect. Artificial intelligence, deepfake, is ushering in an era in which it will be important to maintain our own reality.

Keywords: online emotions, narcissism, reality, perception

Kárpáti, Andrea: The effects of deepfake on education

Deepfake is present at school, but it does not seem to endanger the validity and reliability of the content of education yet. This chapter begins with examples of deepfake that may pose a threat for the world of learning. "What can the teacher do?" – the second part of the chapter begins with this question. We briefly summarise some educational research projects on the prevalence and effects of deepfake. We show educational methods and administrative actions that schools use to prevent the integration of fake knowledge in the minds of children and youth. We also present some good examples about openly acknowledged fakes that employ synthetic imaging techniques based on authentic facts and data. "What can parents do?" – we ask at the end of the chapter. To act efficiently, digital literacy skills of adults need to be developed. Parents must acknowledge the principles of ethical internet use and image copyright regulations. These all are being taught to the children – the question is, how far parents are able and willing to acquire them?

Keywords: authentic digital content, digital content control, digital literacy, media competence, media ethics